A Real-World Evidence Study of Epigenomic Subtyping in Liquid Biopsy–Based Lung Cancer Cohorts

Authors: Jayati Saha[1], Shaun Forbes[1], Sheila Solomon[1], Nicole Zhang[1]

Affiliations: [1]Guardant Health, Redwood City, CA

Introduction

Lung cancer is a heterogeneous disease comprising multiple histologic subtypes with distinct biological and clinical characteristics. Accurate subtype classification is critical for treatment selection and outcomes assessment. However, in real-world data (RWD), histology information is often incomplete or inconsistently captured.

Traditional tissue-based subtyping may not fully reflect tumor heterogeneity, particularly in advanced disease where dynamic molecular changes occur. Emerging approaches, such as **methylation-based molecular subtyping (MLS)**, leverage circulating tumor DNA methylation signals to better characterize tumor biology beyond conventional histology.[1]

This study aimed to evaluate subtype availability, distribution, and molecular characteristics in a large real-world lung cancer baseline cohort, and to assess the potential value of MLS in improving subtype resolution.

Methods

Study Design and Data Source:
This retrospective observational study utilized the InfinityAI real-world database, comprising clinical and molecular data from patients with lung cancer in the United States.

Study Population:
Patients with lung cancer (N=201,543) were identified. Those with available subtype information (N=11,461) were included for further evaluation. A final baseline cohort (N=8,559) was defined based on data completeness criteria.

Subtype Classification:
Patients were categorized into lung cancer subtypes using application-based classification thresholds:

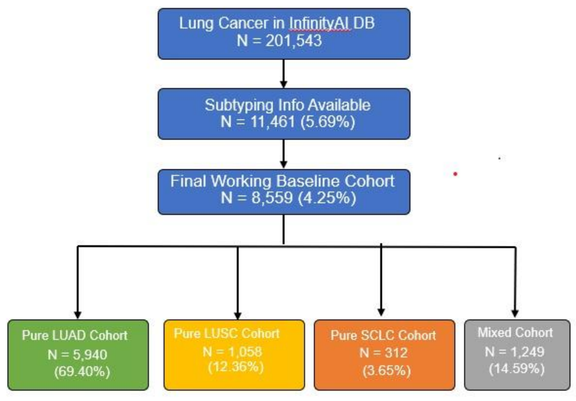
- **Pure Cohort (LUAD, LUSC, SCLC):** ≥90% subtype-specific classification
- **Mixed cohort:** <90% across all subtypes

Outcomes and Analysis:
Descriptive analyses were conducted to evaluate:

- Subtype distribution across cohorts
- Proportion of mixed histology cases
- Prevalence of subtype-specific genomic alterations, including known driver mutations (e.g., LUAD-associated drivers, TP53/RB1 in SCLC)

The findings were interpreted in the context of methylation-based MLS to assess its potential to better capture tumor biology and improve subtype resolution in RWD settings.

Figure 1. Consort Diagram



Results

Among 201,543 lung cancer patients, only 11,461 (5.7%) had subtype information available, and 8,559 met criteria for the final baseline cohort. MLS classified patients as LUAD (69.4%), LUSC (12.4%), SCLC (3.6%), and Mixed (14.6%).

Distinct subtype-specific genomic patterns were observed: LUAD was enriched for targetable drivers (EGFR, KRAS G12C, MET exon 14), LUSC for FGFR, PIK3CA, and PTEN alterations, and SCLC for canonical TP53/RB1 alterations. The Mixed cohort demonstrated overlapping genomic features across subtypes.

Real-world outcomes varied across subtypes, reflecting clinical heterogeneity. LUAD was enriched for actionable driver mutations (e.g., EGFR, KRAS, MET), consistent with availability of targeted therapies.

In contrast, LUSC and SCLC showed fewer targetable alterations and were dominated by tumor suppressor biology (e.g., TP53, RB1). The Mixed cohort demonstrated overlapping genomic features and variable outcomes, highlighting limitations of histology-based classification and supporting the role of methylation-based MLS.

Table 1. Patient demographic and clinical characteristics

Characteristics	Pure LUAD		Pure LUSC		Pure SCLC		Mixed Cohort	
	N/Mean	%/SD	N/Mean	%/SD	N/Mean	%/SD	N/Mean	%/SD
Total N	5,940		1,058		312		1,249	
Age (years)								
<50	265	4%	22	2%	6	2%	31	2%
50-65	1520	26%	233	22%	83	27%	254	20%
>65	4155	70%	803	76%	223	71%	964	77%
Mean (SD)	69.27	10.54	70.13	9.17	69.18	9.06	71.03	9.36
Median (Range)	70.00	(21-85)	70.00	(18-85)	69.50	(35-85)	72	(24-85)
Gender								
Male	2822	48%	664	63%	142	46%	732	59%
Female	3118	52%	394	37%	170	54%	517	41%
Geographic Location								
Northeast	1346	23%	150	14%	62	20%	243	19%
Midwest	1125	19%	255	24%	72	23%	241	19%
South	2575	43%	526	50%	145	46%	608	49%
West	773	13%	114	11%	29	9%	131	10%
Unknown/Missing	121	2%	13	1%	4	1%	26	2%
Smoking*								
Current/Past Smoker (Ever-Smoker)	1795	30%	436	41%	125	40%	472	38%
***ECI**								
Total comorbidities	3.91	2.95	4.21	3.08	4.11	3.03	4.2	3.03
Elixhauser score	5.52	7.57	6.54	8.08	6.06	7.37	6.27	7.77
Unknown/Missing	708	12%	114	11%	36	12%	139	11%
Race								
Caucasian	3533	59%	696	66%	220	71%	837	67%
African American	653	11%	32	3%	5	2%	135	11%
Asian	333	6%	127	12%	22	7%	37	3%
Others	623	10%	95	9%	31	10%	105	8%
Unknown/Missing	798	13%	108	10%	34	11%	135	11%
Ethnicity								
Non-Hispanic	4213	71%	809	76%	229	73%	939	75%
Hispanic	423	7%	61	6%	24	8%	85	7%
Unknown/Missing	1304	22%	188	18%	59	19%	225	18%

* From claim records 6 months prior to Infinity test

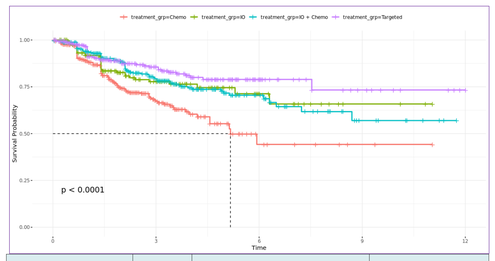
Table 2. Subtype-Specific Key Driver Mutations Across Lung Cancer Cohorts

Biomarker	Pure LUAD	Pure LUSC	Pure SCLC	Mixed Cohort
EGFR (ex19/L858R)	**HIGH**	LOW	--	LOW
KRAS G12C	**HIGH**	LOW	--	LOW
MET exon14	**MOD**	--	--	LOW
FGFR	MOD	**HIGH**	MOD	MOD
PIK3CA	MOD	**HIGH**	MOD	MOD
PTEN	LOW	MOD	MOD	MOD
TP53	**HIGH**	**HIGH**	**HIGH**	**HIGH**
RB1	LOW	MOD	**HIGH**	MOD

High (≥20%) | Moderate (10–19%) | Low (1–9%) | -- (0%)

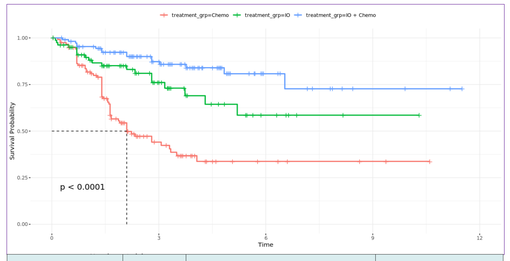
Figure 2. 1L-TTD of Baseline subtypes by Treatment Category

Pure LUAD



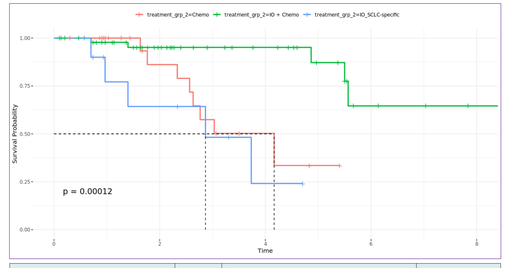
Cohort	N	Median TTD (Month)	95% CL
Chemo	356	5.2	4.6, NR
IO	184	NR	NR, NR
IO+Chemo	362	NR	8.7, NR
Targeted	344	NR	NR, NR

Pure LUSC



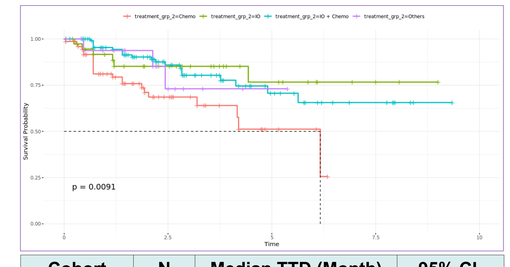
Cohort	N	Median TTD (Month)	95% CL
Chemo	119	2.1	1.6, 3.5
IO	82	NR	5.2, NR
IO+Chemo	112	NR	NR, NR

Pure SCLC



Cohort	N	Median TTD (Month)	95% CL
Chemo	23	4.2	2.6, NR
IO_SCLC_specific	11	2.9	1.4, NR
IO+Chemo	50	NR	5.6, NR

Mixed Cohort



Cohort	N	Median TTD (Month)	95% CL
Chemo	74	6.2	4.2, NR
IO	41	NR	NR, NR
IO+Chemo	121	NR	NR, NR
Others	18	NR	NR, NR

Conclusions

- MLS identifies biologically distinct lung cancer subtypes with expected genomic signatures and differential clinical outcomes in real-world settings.
- A substantial Mixed cohort (14.6%) highlights limitations of traditional histology and underlying tumor heterogeneity.
- MLS has potential to improve subtype resolution and enable more precise patient stratification for real-world and clinical decision-making.

References

1. Guardant Health. Analytical validation of a plasma-based methylation profiling assay for cancer detection and characterization. Presented at: AACR Annual Meeting; 2023.

Questions can be directed to Jayati Saha. jisaha@guardanthealth.com