CLINICAL STUDY DOCUMENT

Statistical Analysis Plan

LUNG-DUAL-MO

Prospective Dual-Cohort Observational Study of Lung Cancer Treatment Response and Tumor Microenvironment Based on Multi-omics Analysis

Unique protocol ID	LCAI0812
Document version	English review draft 1.0
Document date	16 November 2026
ClinicalTrials.gov ID	Not yet assigned; insert after registration, if available.
Source protocol	Prospective dual-cohort observational study registration draft and companion Study Protocol.
Ethics approval number	0978-01
Sponsor	Union Hospital, Tongji Medical College, Huazhong University of Science and Technology
Principal investigator	Heshui Shi, Department of Radiology

Study synopsis

This SAP accompanies protocol LCAI0812. The study is a single-center prospective observational dual-cohort study in 50 adults with lung cancer. Cohort A includes approximately 40 patients with advanced lung cancer receiving clinically selected immunotherapy or immunotherapy-based systemic treatment and uses residual pretreatment biopsy tissue for bulk RNA-seq. Cohort B includes approximately 10 patients with stage IB lung cancer undergoing upfront surgery without neoadjuvant therapy and uses residual fresh surgical tissue for scRNA-seq. ORR, PFS and OS are clinical outcomes for Cohort A; Cohort B is an independent mechanistic cohort.

Document status and provenance

The supplied registration draft contains a cohort-allocation inconsistency: some front-matter fields state approximately 35/15, whereas the revised detailed description and outcomes sections state approximately 40/10. This SAP uses 40/10 as the working operational allocation and requires registry reconciliation before final approval. The analyses below are exploratory because the planned sample is feasibility based and no formal hypothesis-driven power calculation is specified.

1 Scope, design and analysis populations

This SAP prespecifies descriptive, association, survival, transcriptomic, single-cell and cross-cohort analyses for protocol LCAI0812. The study is observational; no treatment, biopsy, surgery or other clinical intervention is assigned by the study.

1.1 Analysis populations and units

Enrolled population	All consented participants meeting cohort eligibility and enrolled prospectively.
Cohort A analysis set	Eligible advanced lung cancer participants with pretreatment clinical data and clinically selected immunotherapy or immunotherapy-based treatment. Omics analyses additionally require assay-quality bulk RNA-seq data.
Cohort B analysis set	Eligible stage IB participants undergoing upfront surgery without neoadjuvant therapy. Single-cell analyses additionally require assay-quality scRNA-seq data.
Primary unit	Participant for clinical outcomes; sample, cell, cell type or molecular feature for specified omics analyses, with participant identity retained as the biological replicate.
Cross-cohort principle	Integrate only prespecified molecular, pathway and/or imaging phenotypes while preserving cohort-specific clinical interpretation.

Cells must not be treated as independent patient-level replicates. Participant-level biological replication should be preserved in differential expression, pathway and cross-cohort inference wherever methodologically applicable.

1.2 Analysis objectives

Cohort A objective	Explore associations of baseline clinical, pathological, imaging and bulk transcriptomic features with objective response, PFS and OS during immunotherapy-based care.
Cohort B objective	Characterize tumor-cell states, immune-cell states, tumor microenvironment composition and cell-cell interactions in stage IB untreated surgical specimens.
Integration objective	Map prespecified signatures or imaging phenotypes arising from Cohort A to cellular states and microenvironmental programs observed in Cohort B.

2 Endpoint derivation and cohort definitions

2.1 Cohort A clinical outcomes

ORR	Proportion of Cohort A participants achieving complete or partial response according to the clinically applied tumor-response criteria documented in routine care. The exact public registry criterion should be verified before naming a specific standard.
PFS	Time from initiation of immunotherapy or immunotherapy-based treatment to documented disease progression or death from any cause, whichever occurs first.
OS	Time from treatment initiation to death from any cause. Participants alive at last known contact are censored on that date.

Follow-up horizon	Planned study follow-up up to approximately 24 months, subject to actual enrollment and study completion dates.

2.2 Cohort B mechanistic outcomes

Cell-state composition	Relative abundance and characteristics of tumor, lymphoid, myeloid, stromal and other identified cell populations, subject to assay quality.
Differential programs	Cell-type or state-specific transcriptional programs and pathway activity.
Cell-cell communication	Exploratory ligand-receptor or other cell-interaction analyses using assay-appropriate methods.
Immunotherapy labels	Not applicable. Cohort B is not categorized as immunotherapy responder/nonresponder because participants do not receive study-observed immunotherapy.

3 Sample size and statistical interpretation

The total planned sample is 50 participants. The revised operational allocation is approximately 40 participants in Cohort A and 10 in Cohort B. This sample size is feasibility based and reflects eligible tissue availability, expected assay success, and the cost/logistical requirements of bulk and single-cell sequencing. No formal hypothesis-driven power calculation is specified.

Accordingly, inferential analyses are exploratory. Effect estimates, confidence intervals and adjusted P values are preferred over dichotomous claims of validation. Multivariable models will be limited by the number of participants and outcome events; model complexity should be reduced when events are sparse.

4 General statistical principles

Statistical tests	Two-sided unless explicitly stated otherwise.
Nominal significance	$P < 0.05$ for conventional exploratory clinical analyses; omics analyses use appropriate multiple-testing correction.
Continuous variables	Mean ± standard deviation when approximately normally distributed; otherwise median and interquartile range.
Categorical variables	Counts and percentages.
Clinical group comparisons	t test or Mann-Whitney U test for continuous variables; chi-square or Fisher exact test for categorical variables, as appropriate.
Confidence intervals	95% confidence intervals where estimable.
Software	Statistical and omics software versions will be recorded in the analysis archive.

5 Cohort A clinical analysis

5.1 Objective response

ORR will be summarized as the number and percentage of evaluable Cohort A participants with complete or partial response, with a 95% confidence interval when appropriate. Baseline clinical, pathological, imaging and molecular features may be compared between response categories. Logistic regression may be used to explore associations with objective response when the number of response events and non-events is sufficient. With sparse events, analyses will be limited to univariable or parsimonious models.

5.2 Progression-free and overall survival

Kaplan-Meier methods will estimate PFS and OS distributions. Median survival and landmark survival estimates may be reported with 95% confidence intervals when estimable. Exploratory group comparisons may use the log-rank test. Cox proportional hazards regression may assess associations of prespecified clinical, imaging or molecular features with PFS or OS when event counts permit. Proportional-hazards assumptions should be checked graphically and/or with standard residual-based procedures.

5.3 Covariate handling

Candidate covariates may include age, sex, smoking history, pathological type, clinical stage, clinically selected treatment regimen, baseline laboratory and imaging variables, and prespecified molecular signatures. The small sample size requires restraint: the final number of covariates should reflect the available event count, and data-driven overfitting should be avoided.

6 Bulk RNA-seq analysis in Cohort A

- Apply assay-appropriate sample-level and feature-level quality control before inferential analysis.
- Normalize expression data using a method appropriate to the sequencing and count-processing pipeline.
- Perform exploratory differential expression analyses according to clinically meaningful response or outcome groups when sample size permits.
- Use false-discovery-rate control or another appropriate multiple-testing procedure for large-scale gene-level testing.
- Perform pathway, gene-set and/or signature analyses to identify biologically coherent programs associated with response, PFS or OS.
- When molecular scores are related to time-to-event outcomes, use continuous-score analyses where feasible rather than outcome-selected cut points.
- Any cut points used for visualization should not be presented as validated clinical thresholds unless independently justified.

7 scRNA-seq analysis in Cohort B

- Apply cell- and sample-level quality control using prespecified thresholds appropriate to the sequencing platform and laboratory pipeline.
- Perform normalization, dimensionality reduction, clustering and cell-type/state annotation using documented methods and reference markers.
- Characterize tumor, lymphoid, myeloid, stromal and other relevant cell populations and their transcriptional programs.
- Assess differential cell-state programs and pathway activity while retaining participant as the biological replicate where possible.
- Explore cell-cell communication using an assay-appropriate ligand-receptor or related framework.
- Adjacent non-tumor tissue, when available, may be used as a within-participant contextual comparator; availability and pairing must be reported.
- Do not label Cohort B participants as immunotherapy responders or nonresponders.

8 Cross-cohort integration

Cross-cohort integration is exploratory and mechanistic rather than a direct validation of treatment response. Gene sets, pathways, bulk-derived signatures and/or imaging phenotypes identified or prespecified in Cohort A may be projected onto Cohort B single-cell data to determine which cell types or cell states express or enrich those programs. The distinct disease stages, tissue acquisition methods and clinical contexts of the two cohorts must be retained in interpretation.

Potential integration strategies include gene-set scoring, pathway enrichment, mapping of bulk signatures to annotated cell populations, comparison of immune or stromal programs, and evaluation of concordance between imaging

phenotypes and cellular states. Any integrative model should be limited in complexity and clearly labeled hypothesis-generating.

9 Missing data and assay failure

Clinical baseline data	Report the amount and pattern of missingness. Complete-case analysis may be used for simple descriptive or exploratory models when missingness is limited; no favorable imputation strategy is to be selected after examining results.
Response follow-up	Document participants without evaluable response and reasons. ORR denominators must be explicitly defined in each analysis.
PFS/OS follow-up	Use the last verified disease assessment or known-alive date according to the endpoint definition; do not invent event dates.
Bulk RNA-seq failure	Participants without assay-quality bulk RNA-seq remain part of the clinical cohort but are excluded from bulk-dependent analyses; reasons for failure are summarized.
scRNA-seq failure	Participants without assay-quality scRNA-seq remain part of the enrolled surgical cohort but are excluded from cell-level analyses; reasons for failure are summarized.
Missing tissue subtype	Unavailable adjacent non-tumor tissue is not imputed and is reported as unavailable.

Because the total sample is small, extensive multiple imputation of high-dimensional molecular data is not planned. Missing clinical covariates may be handled using simple prespecified methods only if justified by the amount and mechanism of missingness.

10 Multiplicity and interpretation

There is no confirmatory hypothesis-testing hierarchy. Clinical association analyses are exploratory. High-dimensional bulk RNA-seq and scRNA-seq analyses will use assay-appropriate multiple-testing correction, typically false-discovery-rate control where applicable. Nominal P values from small subgroup or exploratory clinical analyses should not be interpreted as independent validation.

The study is intended to generate candidate biomarkers and mechanistic hypotheses for later external validation. Results should emphasize effect sizes, uncertainty, biological coherence and reproducibility across analytical levels rather than isolated significance.

11 Sensitivity and robustness analyses

- Repeat key Cohort A clinical analyses after excluding participants with major protocol deviations or inadequate follow-up, clearly labeling these as supportive analyses.
- Where response criteria or assessment timing vary in routine care, describe heterogeneity and perform a sensitivity analysis using the most consistently available assessment definition when feasible.
- For molecular signatures, compare continuous-score results with prespecified categorical summaries used only for visualization or biological interpretation.
- For scRNA-seq, examine whether major conclusions persist after alternative reasonable QC thresholds or participant-level aggregation methods.
- For cross-cohort mapping, report whether findings depend strongly on a single participant, single cell population or single computational method.

12 Data quality, reproducibility and reporting

Linked records	Participant ID, cohort assignment, consent date, sample type, collection date, assay batch, clinical/pathological/imaging variables, treatment dates, response assessments, progression dates and vital status as applicable.
Sample traceability	Maintain coded identifiers linking residual tissue, derived molecular data and participant-level clinical records.
QC records	Archive bulk RNA-seq and scRNA-seq quality-control metrics, exclusion reasons and final analysis datasets.
Analysis archive	Retain software/package versions, scripts, parameter settings, random seeds when used, and final tables/figures.
Reporting	Participant flow, assay success, missingness and reasons for exclusion from each molecular analysis must be reported transparently.

13 Planned presentation

Participant flow	Screened, enrolled, cohort allocation, tissue availability, assay success and clinical follow-up.
Baseline table	Clinical, pathological and imaging characteristics by cohort; Cohort A additionally by clinically relevant response groups when appropriate.
Clinical outcomes	ORR, Kaplan-Meier PFS and OS, with exploratory effect estimates for prespecified features.
Bulk RNA-seq	QC summary, differential expression, pathway/signature analyses and associations with clinical outcomes.
scRNA-seq	Cell atlas/annotation, cell-state programs, immune microenvironment composition and cell-cell communication.
Cross-cohort integration	Mapping of Cohort A molecular/imaging phenotypes to Cohort B cell states, explicitly labeled exploratory.

14 Prespecification and approval

This SAP should be reviewed and approved before final database lock and before outcome-driven selection of analytical strategies. Any analyses added after inspection of outcome-associated results should be identified as post hoc. The final approved version should reconcile the cohort-allocation discrepancy in the registration draft and record the actual investigator/statistical approval date.