PUBLIC-DATA RESEARCH EXAMPLE / 01

TCGA-LUAD + GSE72094

Clinical and molecular data harmonization • Prognostic model development • Exploratory external validation

Development dataset	Independent external dataset	Research conclusion
333 patients / 135 deaths	393 patients / 111 deaths	No stable incremental prognostic benefit from the prespecified molecular panel

1 Research question

Do selected genomics or protein markers improve lung adenocarcinoma survival prediction beyond a clinical baseline in the available public cohorts? This is a prognostic question, not a treatment-effect question.

2 Data readiness & standardization

Processed public TCGA-LUAD clinical metadata and aligned available molecular measurements to unique patient-level records. Applied pre-specified sample eligibility and patient/sample matching. Recorded missingness and restrictions of mutation/protein calls.

Input reviewed	Research-ready output
Clinical / diagnosis / outcome records	Patient-level cohort, age, sex, stage, smoking and survival fields
DNA mutation + RPPA measurements	Four prespecified DNA indicators; five measured protein features, joined to patient IDs
External GSE72094 dataset	393-person external matrix suitable for clinical-only and clinical+DNA comparison

3 Development and validation

Four prespecified Ridge Cox models shared the same 333 patients. Five-fold cross-validation repeated five times formed the outer evaluation, with three-fold inner tuning. Only compatible clinical-only and clinical+DNA models were tested externally; no external re-fitting.

Metric	Clinical (M1)	Clinical + DNA (M2)
Internal validated C-index	0.6551	0.6465
External C-index	0.6104	0.6265
External 3-year Brier (lower better)	0.2194	0.2271

4 Decision: do not claim incremental benefit

External M2 C-index increased by 0.0161 relative to M1 but its paired uncertainty interval crossed zero and 3-year Brier score worsened. The evidence does not support claiming a clinically reliable molecular prognostic improvement. Preserve the documented workflow; pause this specific panel rather than searching post hoc for a favorable result.

Important limitations

Historical cohorts; external mutation assay definitions and time origin not fully aligned; no compatible external RPPA panel; outcomes do not establish treatment benefit. This example is not a clinically validated product or patient-care tool.

Source: TCGA PanCanAtlas (gdc.cancer.gov/about-data/publications/pancanatlas); GEO accession GSE72094 (ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72094). Results based on an October 5, 2026 exploratory report. Not endorsed by TCGA, NCI, NIH or GEO.