

INTRODUCTION

- Squamous Non Small Cell Lung Cancer (SqNSCLC) is a heterogeneous disease linked to poor clinical outcomes. Previous efforts to characterize the molecular landscape of SqNSCLC have used small sample sizes and single data modalities.
- To enable precise therapeutic strategies, a multimodal AI framework was employed to integrate gene expression, copy number alterations, and somatic mutation data to categorize SqNSCLC into distinct molecular subtypes.
- Analysis of these subtypes yielded clinically actionable vulnerabilities and patient stratification strategies critical for advancing the next-generation of lung cancer treatments.

METHODS

1. Tempus real-world database

- De-identified clinico-genomic records from SqNSCLC patients profiled with matching de-identified Tempus xT (DNA) and xR (RNA)-seq assays were analyzed (n = 4,973).

2. High-risk subtype identification

- A multimodal autoencoder architecture was developed to integrate gene expression, CNV, and mutation profiles from SqNSCLC tumors.
- K-means clustering was applied to joint embeddings to define patient subtypes, which were then functionally characterized by molecular enrichment.
- Real-world overall survival (wOS) analysis was performed to assess the clinical and prognostic relevance of the identified subtypes.

3. Spatial niche identification

- SingleR was used to assign cell types using reference marker genes from Salcher *et al.*, 2022.
- Spatial niches were identified using CellCharter (Varrone *et al.*, 2023).
- Cell-cell communication analysis was performed using LIANA (Dimitrov *et al.*, 2024).

ACKNOWLEDGMENTS

We thank Vajsa Shah and ChiShing Ho from the Tempus spatial team for providing spatial niches and Amrita A. Iyer, Ph.D, from the Tempus Science Communications team for poster review

RESULTS

Figure 1. Multimodal AI identifies seven SqNSCLC subtypes with significant differences in overall survival

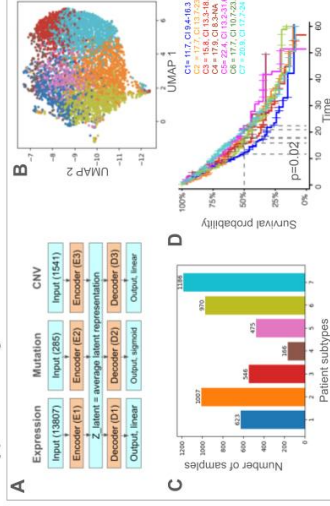


Figure 1. (A) Multimodal autoencoder architecture using expression, mutation, and CNV data of SqNSCLC patients. (B) K-means clustering was performed on the joint embeddings. Clustering metrics (Dunn index) identified K=7 as optimal. UMAP of the joint embeddings representing seven distinct patient clusters and (C) distribution. (D) Kaplan-Meier survival curves showing wOS of seven SqNSCLC subtypes.

Figure 2. NF1 mutations are enriched in subtype associated with worst prognosis

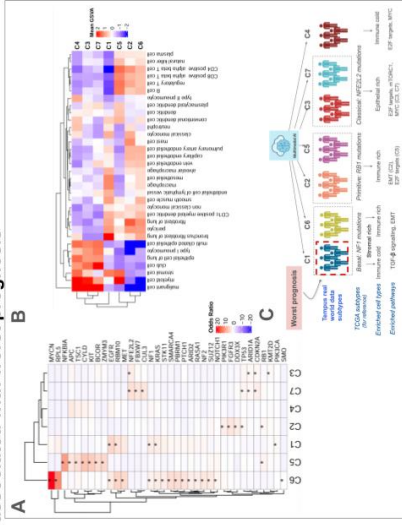


Figure 2. (A) Odds ratio represents enrichment of pathogenic mutations. (B) Heatmap of mean GSVA (Gene Set Variation Analysis) scores of lung cell type signatures derived from Salcher *et al.*, 2022. (C) Summary of SqNSCLC subtypes, key pathogenic mutations, cell types and enriched pathways.

SUMMARY

- Multimodal AI derived SqNSCLC subtypes differ in prognosis, and reveal additional intra-subtype heterogeneity.
- A robust 31-gene C1 signature and spatial dissection of the tumor microenvironment identified novel and known fibroblast-tumor signaling axes.
- This study validated the potential of multimodal omic integration for high-resolution patient subtyping, establishing a critical foundation for developing integrative AI frameworks to accelerate precision oncology.

Figure 3. 31 gene C1 signature is associated with poor survival across cohorts

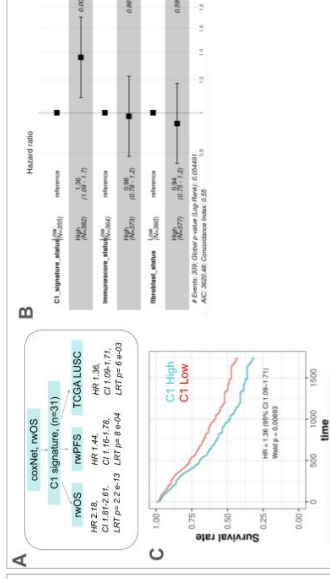


Figure 3. (A) 31 gene C1 signature was derived using CoxNet with wOS, wOS derived 31 gene signature was also predictive of wPFS (real-world Progression Free Survival) and OS in TCGA LUSC Lung Squamous Cell Carcinoma. (B) Multivariate forest plot adjusting for clinical variables along with immune and fibroblast scores (Kocil, Aran *et al.*, 2017) in TCGA LUSC cohort. Indicates p-value <0.05 (C) Kaplan-Meier survival curves showing high C1 signature separates patients with poor outcomes in TCGA LUSC cohort.

Figure 4. Spatial interaction between fibroblasts and malignant cells shapes pro-tumorigenic environments in poor prognosis patients

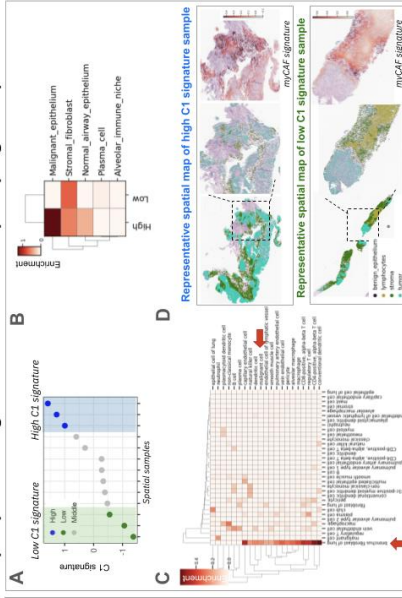


Figure 4. (A) Mapping the 31-gene C1 signature onto samples with 10x Visium HD spatial transcriptomics data. (B) Enrichment of spatial niches in C1 high (n=3) and low patients (n=3). Only significant associations p<0.05 are shown. (C) Differential neighborhood enrichment analysis between C1 high and low patients, only significant associations p<0.05 are shown. (D) Representative spatial maps of high and low C1 signature patients, representing cell type composition and myCAF (Cancer Associated Fibroblasts) signature.

Figure 5. A bronchus fibroblast paracrine signaling axis suggests a novel therapeutic vulnerability in lung cancer

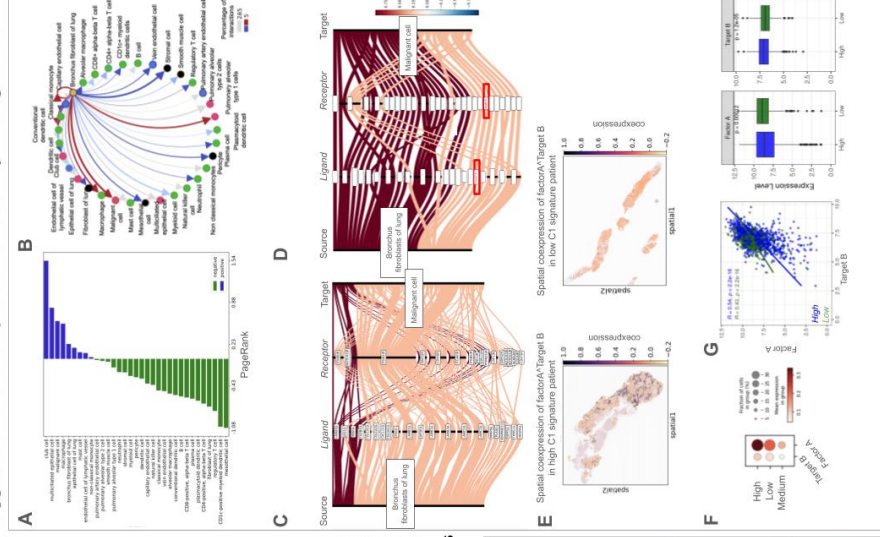


Figure 5. (A) PageRank centrality score representing relative importance of cell type in high vs. low C1 signature patients. (B) Differential cell-cell communication analysis between bronchus fibroblasts with other lung cell types. (C) Known and novel (D) ligand-receptor pairs between bronchus fibroblasts and malignant cells. (E) Spatial co-expression patterns between fibroblast factor A and malignant cell target B in representative C1 high and low patients. (F) Expression of fibroblast factor A and target B in spatial transcriptomics and (G) in real-world bulk transcriptomics data.

ACKNOWLEDGMENTS

We thank Vajsa Shah and ChiShing Ho from the Tempus spatial team for providing spatial niches and Amrita A. Iyer, Ph.D, from the Tempus Science Communications team for poster review

Correspondence: swati.kaushik@tempus.com, radia.johnson@tempus.com