

# Impacts of tobacco smoking on the genomic architecture of *ALK*-rearranged non-small cell lung cancer

Hansol Park<sup>1,2</sup>, Yunah Lee<sup>2</sup>, JungHan Oh<sup>3</sup>, Mi-Sook Lee<sup>3,4</sup>, Ji Young Song<sup>4</sup>, Sung Bin An<sup>3,4</sup>, Yoon-La Choi<sup>3,4</sup>, and Young Seok Ju<sup>1,2</sup>

1. Inocras, Inc. 2. Korea Advanced Institute of Science and Technology (KAIST) 3. Sungkyunkwan University 4. Samsung Medical Center (SMC)

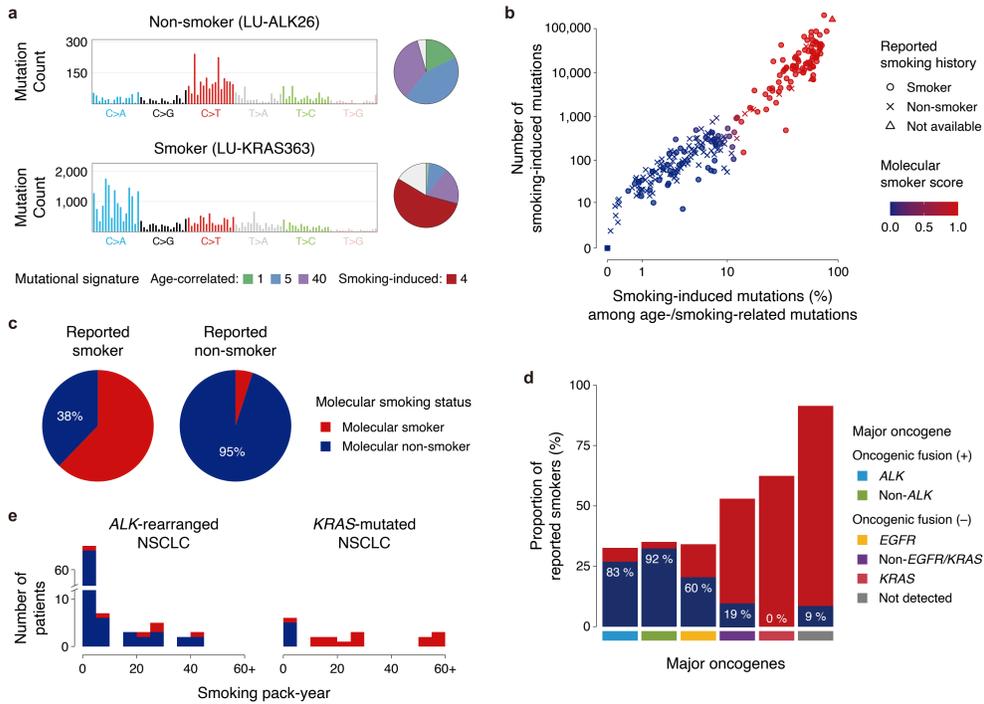
## INTRODUCTION

- ALK* gene fusion is a major oncogenic driver in non-small cell lung cancer (NSCLC), accounting for ~3–7% of cases and most frequently involving *EML4* as the fusion partner.
- Despite its clinical significance, the mechanisms and characteristics of *ALK*-rearranged NSCLCs remain poorly understood.

- ALK*-rearranged NSCLC has been reported primarily in non-smokers and is considered a relatively homogeneous subset of NSCLCs.
- Here, we analyzed whole-genome sequencing data from 89 *ALK*-rearranged NSCLCs alongside 193 NSCLCs harboring other driver mutations, including *RET*, *ROS1*, *EGFR*, and *KRAS*.

## RESULT

### 1. Smoking leaves minimal mutational footprint in *ALK*-rearranged NSCLC



**Figure 1 | Heterogeneous impact of tobacco smoking on genomic mutagenesis across NSCLCs**

**a.** Typical genomic mutation patterns of a non-smoker and a smoker. Smoking-induced mutational signature is exclusively observed in the smoker.

**b.** Clustering of NSCLCs based on the level of smoking-induced mutations. A Gaussian Mixture Model (GMM) identified two distinct molecular subtypes (molecular smokers and non-smokers), where the molecular smoker score is defined as the posterior probability of a patient belonging to the molecular smoker cluster.

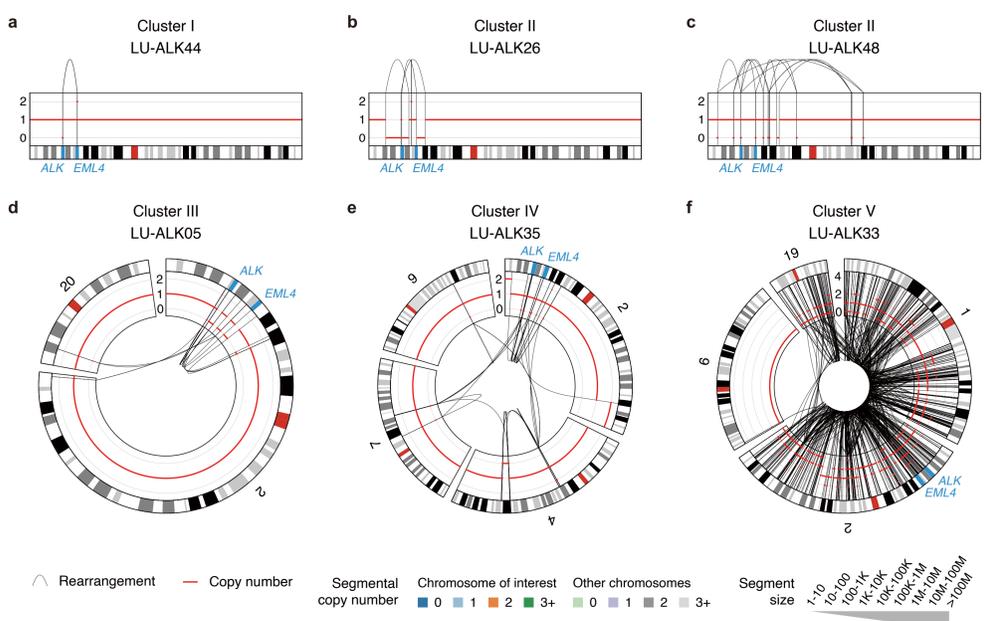
**c.** Proportion of molecular smokers and non-smokers within each clinically reported smoking group. Notably, 38% of clinically reported smokers are classified as molecular non-smokers.

**d.** Proportion of clinically reported smokers and their molecular smoking status. The percentage of molecular non-smokers among clinically reported smokers is shown.

**e.** Smoking pack-years of *ALK*-rearranged and *KRAS*-mutated NSCLCs. Despite high smoking pack-years, certain cases were classified as molecular non-smokers, exhibiting minimal smoking-induced mutations.

### 2. A novel framework identified distinct *ALK* rearrangement clusters

### 3. *ALK* rearrangement clusters vary in genomic complexity

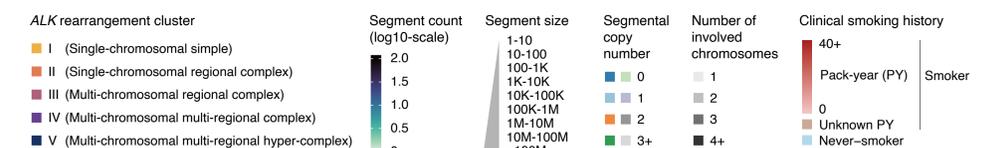
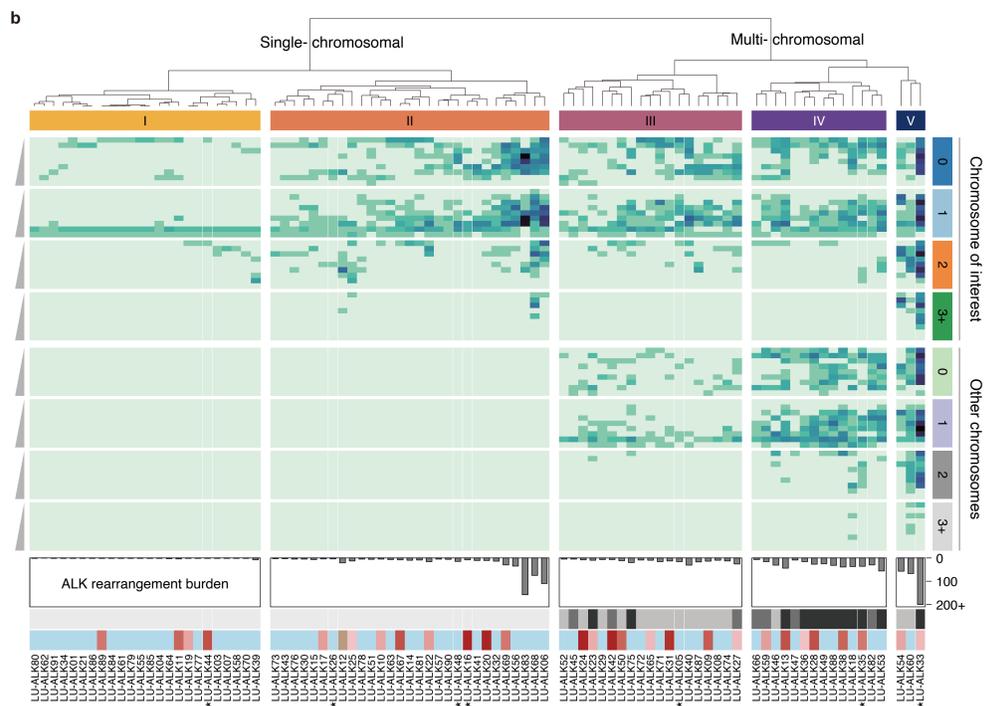
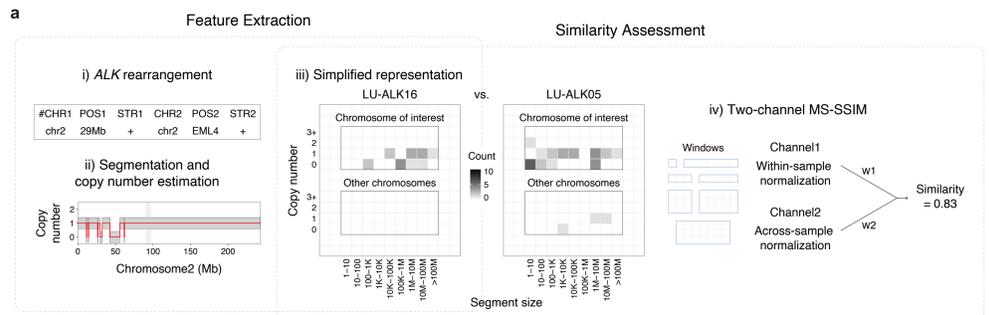


**Figure 3 | Representative examples of each *ALK* rearrangement cluster**

**a.** Cluster I: A reciprocal inversion.

**b-c.** Cluster II: Complex intra-chromosomal rearrangements accompanied by multiple segment losses (b) or minimal segment loss (c), resembling classical and copy-balanced chromothripsis, respectively.

**d-f.** Clusters III-V: Complex multi-chromosomal rearrangements. From Cluster III (d) to Cluster V (f), rearrangement patterns become progressively more complex, involving an increasing number of segments and chromosomes.

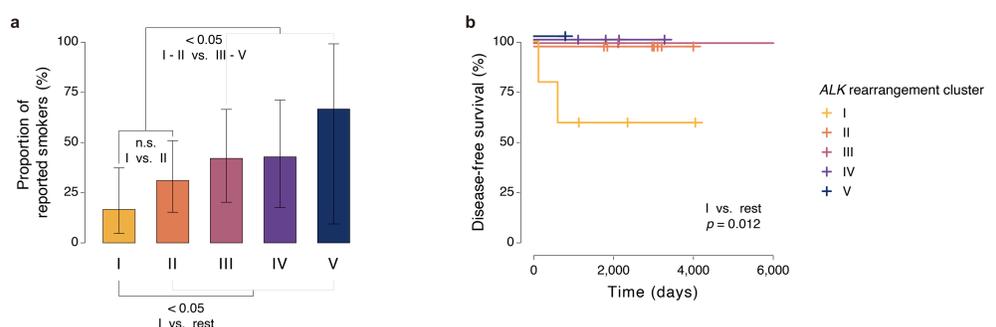


**Figure 2 | Systematic clustering of *ALK* rearrangements**

**a.** Overview of the workflow for assessing similarities among *ALK* rearrangements with the Multi-Scale Structural Similarity Index Measure (MS-SSIM), relying solely on rearrangements and thus independent of preceding or subsequent copy number changes.

**b.** *ALK* rearrangements were classified by the number of involved chromosomes (single vs. multiple) and further subdivided by complexity, resulting in five distinct clusters. Columns represent individuals; rows correspond to segments defined by size and copy number. Cell color (cyan scale) indicates the number of corresponding segments.

### 4. *ALK* rearrangement complexity is associated with smoking and clinical outcome



**Figure 4 | Correlation of *ALK* rearrangement complexity with smoking and clinical outcome**

**a.** Proportion of clinically reported smokers across *ALK* rearrangement clusters. Clinically reported smokers tend to harbor more complex *ALK* rearrangements.

**b.** Patients with simple intra-chromosomal rearrangements (Cluster I) exhibit significantly shorter disease-free survival ( $p = 0.012$ ) compared to the more complex clusters (Clusters II–V).

## DISCUSSION

### Largest cohort

- Due to the low incidence of *ALK*-rearranged NSCLCs, large-scale analyses have been challenging. To our knowledge, this study represents the largest cohort analyzed to date.

### Novel framework for clustering rearrangements

- Our systematic framework for clustering rearrangements provides a new approach to defining rearrangement signatures, with the potential to uncover underlying etiologies and clinically relevant subgroups.

### New perspective on tobacco smoking

- About one third of *ALK*-rearranged NSCLC patients are smokers, yet their tumors harbor few smoking-induced mutations, irrespective of pack-years. It is plausible that they originate from cells largely unaffected by direct tobacco exposure.
- Notably, *ALK* rearrangements were more complex in smokers. While these tumors exhibit minimal smoking-related mutations, suggesting limited direct tobacco effect, smoking-induced immune exhaustion may have impaired immune surveillance, allowing these complex tumors to persist.