4th International Iranian Conference on BIOINFORMATICS
Faculty of Engineering, University of Zanjan
February 4-6 ,2025

ICB13
Iranian Conference on Bioinformatics
سیزدهمین همایش ملی بیوانفورماتیک ایران
دانشکده مهندسی دانشگاه زنجان ـ ۱۶ تا ۱۸ بهمن ۱۴۰۳

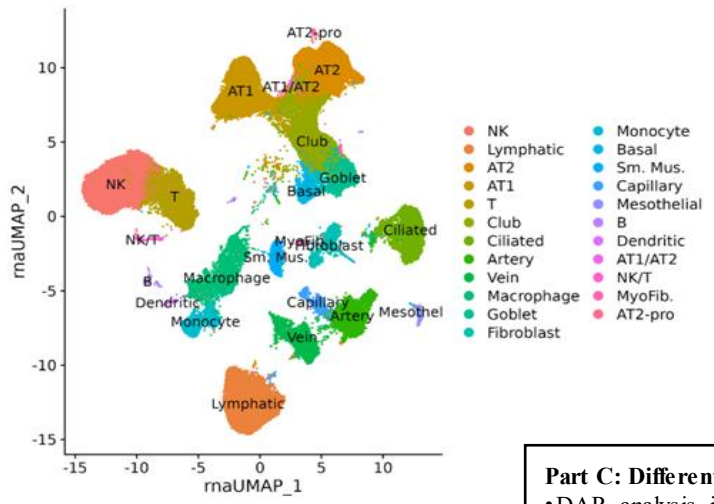
Dissecting the impact of smoking on epigenetic mechanisms that influencing Lung cancer susceptibility

Mahboobeh Golchinpour[1], Zahra Salehi[2], Alireza Fotuhi Siahpirani[1]*

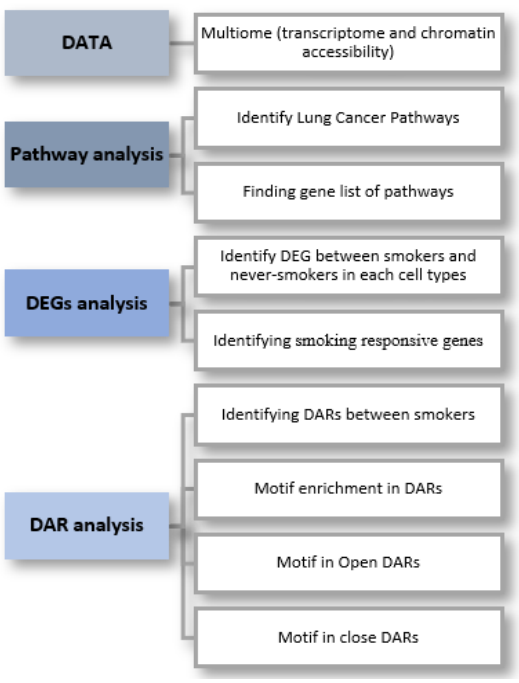
[1] Laboratory of Bioinformatics and Computational Genomics (LBCG), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[2] Hematology, Oncology and Stem Cell Transplantation Research Center, Research Institute for Oncology, Hematology and Cell Therapy, Tehran University of Medical Sciences, Tehran, Iran
* Corresponding Author: a.fotuhi@ut.ac.ir

Introduction
Environmental factors such as smoking is a major risk factor in lung cancer. In this work, we aim to dissect the impact of smoking on the epigenetic mechanisms that influence lung cancer susceptibility.

Data
We used a multiome (transcriptome and chromatin accessibility) maps from 117,911 lung cells with a total of 36,602 genes, derived from 4 males and 4 females in each group of smokers and never-smokers, which examines the epigenetic landscapes and gene regulation in human lung cells. Figure 1 shows the distribution of cell counts across 23 different cell types in the dataset.



Overview of the approach

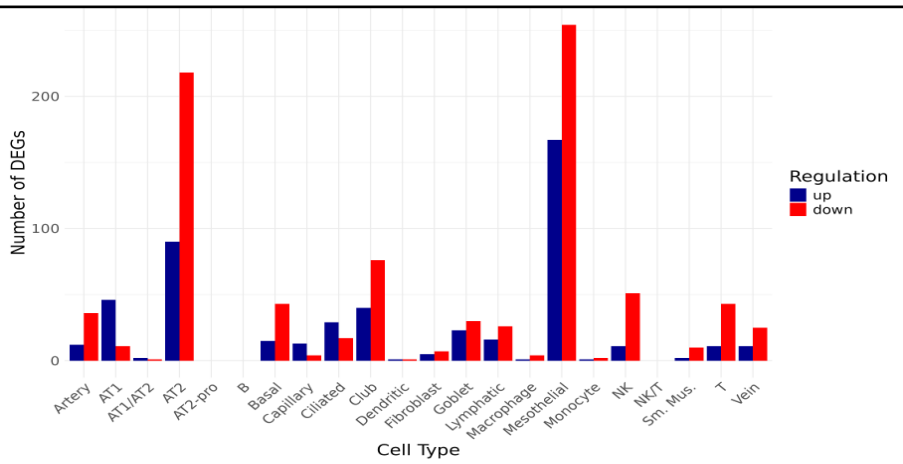


Results

Part A: Pathway analysis
- Analysis of KEGG pathways and research papers identified 2,585 key genes.

Part B: Differentially Expressed Genes (DEGs) analysis
- Differentially expressed genes (DEGs) were determined by comparing gene expression between smokers and non-smokers using the FindMarkers function in Seurat.

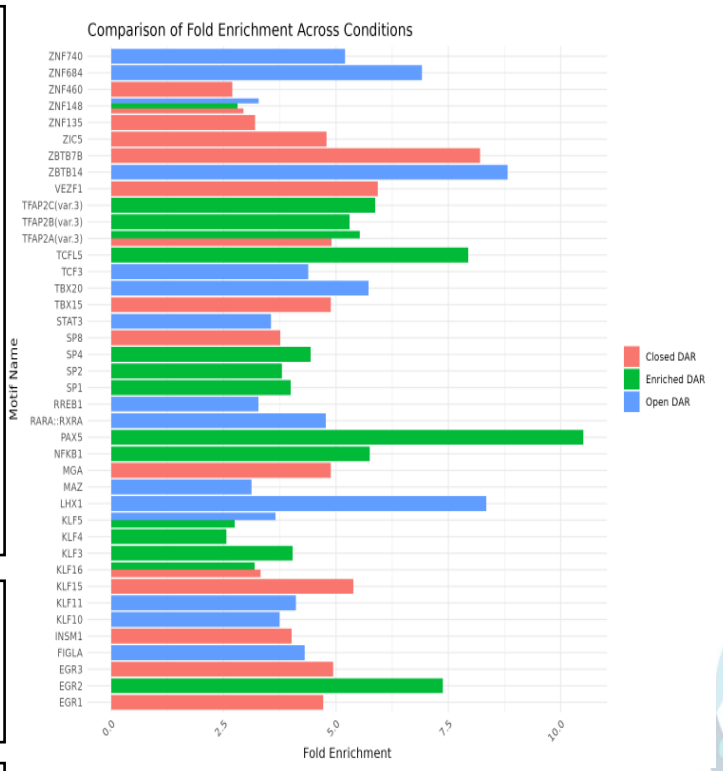
Smoking responsive genes in each cell type
- Identified smoking-responsive genes in 23 cell types and analyzed nearby differentially accessible regions (DARs) to examine chromatin changes. Figure2 show smoking-responsive genes by cell type.



Part C: Differentially Accessible Region (DAR) analysis
- DAR analysis in AT2 cells assessed smoking's effects on chromatin accessibility in lung adenocarcinoma.
- Differential peak analysis compared smokers and non-smokers in AT2 cells using Seurat's FindMarkers() function.

Motif enrichment in DAR
- Motif enrichment analysis was performed on the top differentially accessible peaks of AT2 cell. Top 12 enriched transcription factor binding motifs identified: including SP4, TFAP2C, EGR2, SP1, PAX5, SP2, TFAP2B, ZNF148, KLF5, NF-κB1, and TCFL5.

Motif enrichment in open and closed DARs
- For AT2 DAR, we classified the regions into 'open' and 'closed' DARs based on their accessibility changes.
- Motif enrichment analysis performed in 'open' and 'closed' DARs.
- Figure 3 shows motifs that are enriched across 3 conditions of AT2 cell (Open DAR, Closed DAR, DAR).



Conclusions
The study reveals smoking's impact on chromatin accessibility in AT2 lung cells and identifies key transcription factors, providing insights for targeted therapies.

References
1. Long, et. al. Nat Commun 2024.
2. Abolfathi, et. al. Respir Res 2023.
3. Choudhary, et. al. Cancers (Basel) 2023.
4. Itokawa, et. al. Nat Commun 2022.