



Machine Learning-Driven Proteomic Signature of Smoking Enables Population Reclassification and Reveals Lung Cancer Biomarkers

Jingwen Diao, Xianzhe Fan, Rui Gao, Yihai Wei, Jingdong Zhou, Liheng Shen,
Jianxing He*, Wenhua Liang*

First Affiliated Hospital of Guangzhou Medical University, State Key Laboratory of Respiratory Disease,
Department of Surgery and Oncology, Guangzhou, Guangdong, China



IASLC 2026 World
Conference on Lung Cancer
SEPTEMBER 12 – 15, 2026
SEOUL, REPUBLIC OF KOREA

INTRODUCTION

Rationale and objective

- Self-reported smoking categories incompletely capture molecular exposure.
- We derived an SMF from 2,923 plasma proteins in 52,762 UKB participants.
- We tested lung cancer risk, mediation, PRS interaction, and population reclassification.

METHODS

Study design and analysis

- UK Biobank baseline Olink proteomics cohort.
- Linear regression identified smoking-associated proteins.
- Elastic net with 10-fold CV built the 654-protein SMF.
- Cox, mediation, and PRS interaction models evaluated lung cancer outcomes.

REFERENCES & ACKNOWLEDGEMENTS

Acknowledgements

We thank UK Biobank participants, the UK Biobank team, and the Olink proteomics resource.

Funding

National Key R&D Programme; NSFC; China Postdoctoral Science Foundation; Guangzhou Science and Technology Planning Project; Guangzhou National Laboratory.

Contact

Jingwen Diao: katherined2030@163.com
Correspondence: Jianxing He; Wenhua Liang

DISCUSSION AND CONCLUSION

Interpretation

- SMF captures smoking biology beyond self-report.
- Proteomic alterations mediated nearly half of smoking-related lung cancer risk.
- High-SMF non-smokers suggest exposure mimics; low-SMF smokers suggest molecular resilience.

Clinical implications

- SMF plus PRS may help prioritize low-dose CT screening after validation.

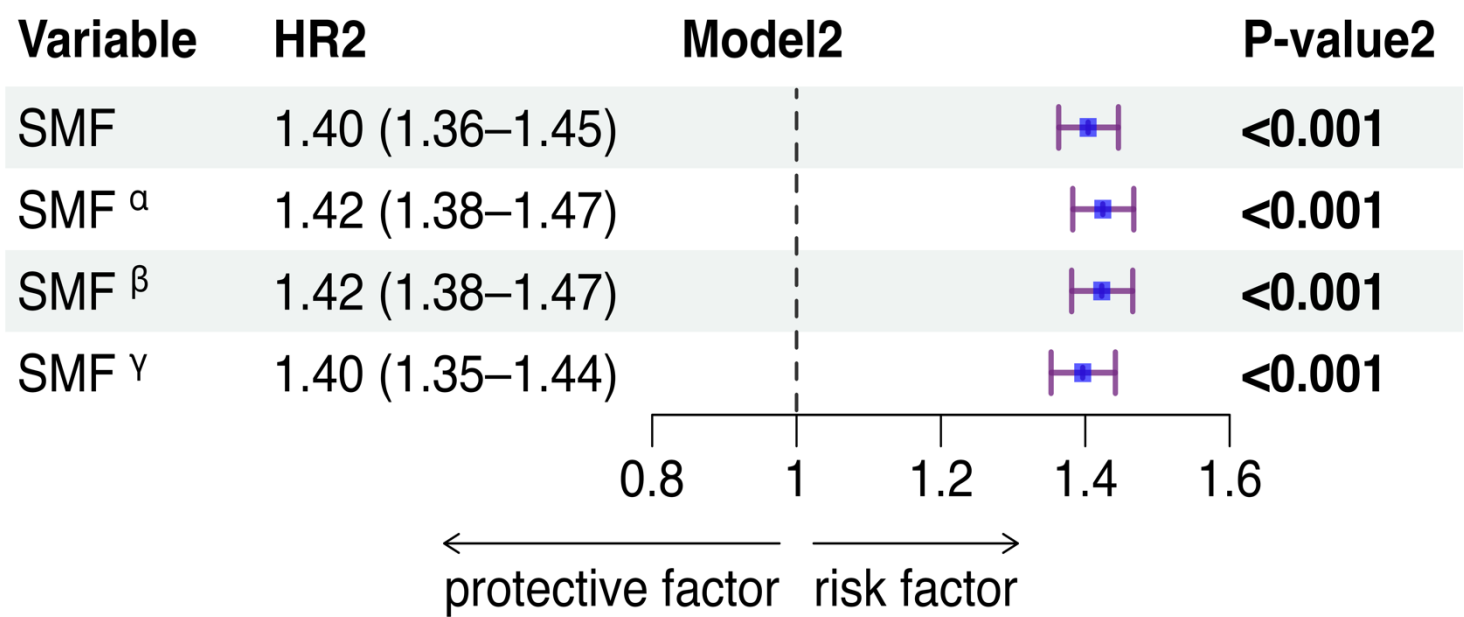
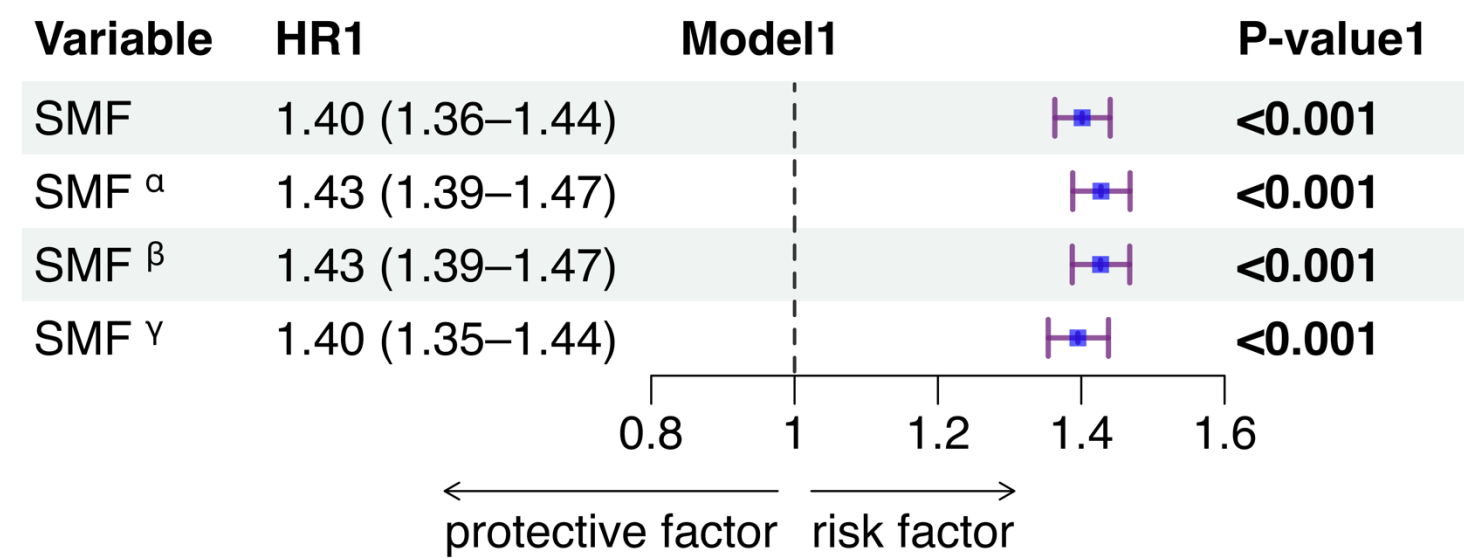


Figure 2. SMF remained associated with lung cancer incidence, lung cancer mortality, and all-cause mortality after adjustment.

RESULTS

Key results

- 1,312 smoking-associated proteins; 654-protein SMF explained 63.6% of smoking-status variance.
- Each 1-SD higher SMF predicted lung cancer incidence/mortality (HR 1.40, 95% CI 1.35-1.44) and all-cause mortality (HR 1.10, 95% CI 1.09-1.11).
- Current smoking remained strongly associated with incident lung cancer (HR 21.58) and lung cancer mortality (HR 20.45).
- SMF mediated 47.3% of smoking's total effect; high PRS plus high SMF marked the highest-risk subgroup.
- Reclassification: high SMF in 41.4% of non-smokers and low SMF in 6.1% of smokers.

Figure 1. Study workflow, proteomic feature construction, and outcome modeling framework.

