Stratifying Lung Adenocarcinoma Risk with Multi-ancestry Polygenic Risk Scores in East Asian Never-Smokers

Batel Blechter†,1, Xiaoyu Wang†,1,2, Jianxin Shi†,1, Kouya Shiraishi †,3, Jiyeon Choi †,1, Keitaro Matsuo †,4, Tzu-Yu Chen †,5,6, Juncheng Dai †,7,8, Rayjean J Hung †,9, Kexin Chen †,10, Xiao-Ou Shu †,11, Young Tae Kim12,13, Parichoy Pal Choudhury14, Jacob Williams1, Maria Teresa Landi1, Dongxin Lin15, Wei Zheng11, Zhihua Yin16, Bao Song17, I-Shou Chang18, Yun-Chul Hong19, Never Smoker Lung Cancer Working Group, Nilanjan Chatterjee ‡,20,21, Olga Y Gorlova ‡,22,23, Christopher I Amos ‡,22,23, Hongbing Shen ‡,7,8, Chao Agnes Hsiung ‡,5, Stephen J Chanock ‡,1, Nathaniel Rothman ‡,9, Takashi Kohno ‡,3, Qing Lan ‡,1, Haoyu Zhang ‡,1

1Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA
2Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc, Rockville, MD, USA
3Division of Genome Biology, National Cancer Research Institute, Tokyo, Japan
4Division of Molecular Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan
5Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan
6Department of Applied Mathematics, Tunghai University, Taichung, Taiwan
7Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China
8Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China
9Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada
10Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China
11Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center and Vanderbilt-Ingram Cancer Center, Nashville, TN, USA
12Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea
13Cancer Research Institute, Seoul National University College of Medicine, Seoul National University
14Department of Epidemiology, Biostatistics, and Prevention Sciences, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
15Division of Cancer Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
16Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
American Cancer Society, Atlanta, GA, USA

Department of Etiology & Carcinogenesis and State Key Laboratory of Molecular Oncology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Department of Epidemiology, School of Public Health, China Medical University, Shenyang, China

Department of Oncology, Shandong Cancer Hospital and Institute, Shandong Academy of Medical Sciences, Jinan, China

National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan

Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Department of Medicine, Section of Epidemiology and Population Science, Institute for Clinical and Translational Research, Houston, TX, USA

Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA

† These authors contributed equally to this work

‡ These authors jointly directed this work

Correspondence to: Batel Blechter (batel.blechter@nih.gov), Qing Lan (qingl@mail.nih.gov), and Haoyu Zhang (haoyu.zhang2@nih.gov)
Abstract

Background
Lung adenocarcinoma (LUAD) among never-smokers is a public health burden especially prevalent in East Asian (EAS) women. Polygenic risk scores (PRSs), which quantify genetic susceptibility, are promising for stratifying risk, yet have mainly been developed in European (EUR) populations. We developed and validated single- and multi-ancestry PRSs for LUAD in EAS never-smokers, using the largest available genome-wide association study (GWAS) dataset.

Methods
We used GWAS summary statistics from both EAS (8,002 cases; 20,782 controls) and EUR (2,058 cases; 5,575 controls) populations, as well as independent EAS individual level data. We evaluated several PRSs approaches: a single-ancestry PRS using 25 variants that reached genome-wide significance (PRS-25), a genome-wide Bayesian based approach (LDpred2), and a multi-ancestry approach that models genetic correlations across ancestries (CT-SLEB). PRS performance was evaluated based on the association with LUAD and AUC values. We then estimated the lifetime absolute risk of LUAD (age 30-80) and projected the AUC at different sample sizes using EAS-derived effect-size distribution and heritability estimates.

Findings
The CT-SLEB PRS showed a strong association with LUAD risk (odds ratio=1.71, 95% confidence interval (CI): 1.61, 1.82) with an AUC of 0.640 (95% CI: 0.629, 0.653). Individuals in the 95th percentile of the PRS had an estimated 6.69% lifetime absolute risk of LUAD. Comparison of LUAD risk between individuals in the highest and lowest 20% PRS quantiles revealed a 3.92-fold increase. Projection analyses indicated that achieving an AUC of 0.70, which approaches the maximized prediction potential of the PRS given the estimated genetic variance, would require a future study encompassing 55,000 EAS LUAD cases with a 1:10 case-control ratio.

Interpretations
Our study underscores the potential of multi-ancestry PRS approaches to enhance LUAD risk stratification in never-smokers, particularly in EAS populations, and highlights the necessary scale of future research to uncover the genetic underpinnings of LUAD.

**Key Words:** Polygenic risk scores, Lung adenocarcinoma, Genome-wide association studies, East Asian never smokers, Lifetime absolute risk
Introduction

Lung cancer is a major global health challenge, responsible for about 20% of all cancer deaths in 2020. While smoking is the primary etiologic factor, around 25% of lung cancer cases occur in never-smokers, with significant geographical variations. Notably, in East Asia (EAS), never-smoking women exhibit high incidences of lung adenocarcinoma (LUAD), the predominant histologic subtype.

Beyond smoking, risk factors for LUAD include environmental and occupational exposures, lifestyle, family history, and genetic susceptibility. Genome-wide association studies (GWAS) for never-smoking lung cancer across EAS and European (EUR) populations have identified 28 susceptibility variants at 25 independent loci. These GWAS findings enhance our understanding of LUAD's genetic architecture, yet the translation of these findings into clinical practice requires further investigation.

Polygenic risk scores (PRS) aggregate the effects of individual single nucleotide polymorphisms (SNPs) to estimate genetic risk for diseases, playing a crucial role in cancer prevention and screening. Previous efforts primarily utilized EUR populations for lung cancer PRS development to better identify high-risk individuals. Nonetheless such PRSs often fall short for EAS never-smokers, reflecting a bias in screening guidelines and the focus of studies on populations with EUR ancestry. Despite ongoing efforts to integrate PRS into clinical practices for complex diseases, as seen in initiatives like the electronic Medical Records and Genomics (eMERGE) network, Veterans Affairs Genomic Medicine at Veteran Affairs (GenoVA) study, and the Women Informed to Screen Depending On Measures of risk (WISDOM) study, lung cancer has been excluded. The predominant global burden of lung cancer, driven primarily by smoking, overlooks the significant impact and potential benefits of PRS in never-smoking populations.
The recent expansion of GWAS across diverse populations, combined with advancements in PRS methodology, underscores a pivotal shift towards enhancing disease outcome prediction beyond EUR populations\textsuperscript{19}. Notably, the development of multi-ancestry PRS methods marks a considerable advancement, leveraging data from various ancestral backgrounds to enrich the predictive accuracy and robustness of PRSs\textsuperscript{20–26}. By jointly modeling genetic data from multiple populations, these approaches improve the predictive power of PRSs, thereby enhancing disease outcome predictions in non-EUR populations.

While the development and validation of PRSs are critical, the application of an established PRS in estimating the absolute risk of a disease offers valuable insights into risk stratification, potentially guiding clinical interventions, such as lung cancer screening. Moreover, projecting necessary sample sizes for future studies to achieve varying levels of PRS predictive accuracy is an underexplored area that could significantly impact epidemiological research design and cost-effectiveness.

This study aims to evaluate the predictive performance of single- and multi-ancestry PRSs for LUAD in never-smoking EAS individuals using state-of-the-art methodologies. By integrating multiple datasets, we constructed PRSs based on summary statistics for 5,622 never-smoking cases and 21,813 never-smoking controls from EAS and EUR ancestries. We assessed the performance of the PRSs using independent, individual level data of 4,438 never-smoking EAS cases and 4,544 never-smoking EAS controls. Furthermore, we estimated the lifetime and 10-year absolute risks of LUAD using the most accurate PRSs developed. Lastly, we projected the sample sizes needed in future research to achieve specific levels of prediction accuracy with PRSs in EAS never-smokers (Figure 1).
Methods

Study design and data sources

Training data compilation: study population and genotyping

The studies, genotyping protocols, and quality control for PRS construction are previously described in detail. Briefly, EAS single-ancestry PRSs for LUAD were constructed using summary data from 3,564 never-smoking LUAD cases and 16,238 never-smoking controls of EAS ancestry from the Nanjing Lung Cancer Study (NJLCS), National Cancer Center of JAPAN (NCC), and the Research Institute and Aichi Cancer Center (ACC). The NJLCS study combined data from several cities, genotyped by Affymetrix Genome-Wide Human SNP Array 6.0, as well as the Oncoarray GWAS. The NCC study included lung cancer patients and control data from multiple Japanese, with genotyping conducted on Illumina HumanOmniExpress and HumanOmni1-Quad genotyping platforms. Similarly, the ACC study pooled data from the multiple Japanese medical institutions, and the Nagahama Study, with genotyping performed using Illumina 610k and Illumina660k platforms. To construct the multi-ancestry PRS, we further incorporated GWAS summary statistics from 2,058 never-smoking LUAD cases and 5,575 never-smoking controls of EUR ancestry, genotyped using Illumina Infinium OmniExpress-24 v1.2 BeadChips and Illumina Human660W-Quad BeadChip.

Tuning and validation: study population and genotyping

For PRS tuning and validation, we used 4,438 never-smoking LUAD cases and 4,544 never-smoking controls from the Female Lung Cancer Consortium in Asia (FLCCA), an international consortium composed of never-smoking EAS women from regions including Mainland China, Singapore, Taiwan, Hong Kong, South Korea and Japan. All samples were genotyped using the Illumina 660W, 370K or 610Q microarrays. The FLCCA data were randomly and equally divided for tuning (2,219 cases, 2,272 controls) and validation (2,219 cases, 2,272 controls). Informed consent was obtained from all participants, and the recruitment and data collection
procedures were approved by the local ethics review committees of all participating research institutes.

**Construction of PRS**

**Single-ancestry PRS methods**

**PRS-25**: We constructed a simple PRS using the 25 independent SNPs (PRS-25) that reached genome-wide significance (i.e., $P<5\times10^{-8}$) in the largest GWAS for LUAD in EAS population to date. The PRS-25 was calculated by summing the risk alleles weighed by their effect sizes (i.e., per allele log-odds ratio) obtained from a meta-analysis including only never-smokers.

**PRS-CT**: We generated a PRS using the CT method using PLINK 1.90. This involved the clumping of SNPs based on LD with a threshold of $r^2=0.1$ within a 500kb window. The LD were estimated from half of the FLCCA dataset used for tuning dataset. We created nine SNP subsets by applying incremental P-value thresholds (i.e., $P<5\times10^{-8}$, $P<5\times10^{-7}$, ..., $P<5\times10^{-1}$, 1), and calculated their respective PRSs using PLINK2 command “--score cols=+scoresums,-scoreavgs no-mean-imputation”. Using the tuning dataset, the AUC was calculated for each threshold to identify the most predictive P-value cutoff.

**LDpred2-PRS**: We applied the LDpred2 method, implemented in the R package bigsnpr, using a Bayesian framework to estimate SNP effect sizes through a shrinkage estimator. This method leverages GWAS summary statistics, incorporating a prior for effect sizes, while also accounting for LD across SNPs. Our analysis was constrained to HapMap3 variants, and we calibrated the model using a range of hyperparameters: the proportion of causal SNPs was set across a 21-point logarithmic sequence from $10^{-5}$ to 1, and the per-SNP heritability was set as a fraction (0.3, 0.7, 1 or 1.4) of the total heritability estimated by LD score regression divided by the number of causal SNPs. Lastly, we used the “sparse” option to set weaker effects to zero. The optimal tuning parameters were selected based on the highest AUC achieved on the tuning dataset.
We evaluated the performance of a European-derived PRS (PRS-EUR 128), which consists of 128 variants obtained from a GWAS of predominately active smokers, on EAS populations to assess its cross-population applicability.

Multi-ancestry PRSs methods:

LDpred2 PRS + PRS-EUR 128: For the multi-ancestry PRSs, we applied the weighted-PRS approach by linearly combining the most predictive single-ancestry PRS from the EAS population (Ldpred2 PRS) with the EUR population-specific PRS (PRS-EUR 128). The weights of this combined PRS were calculated by applying a logistic regression on the tuning dataset in R version 4.2.0.

PRS-CSx: We applied PRS-CSx, a multi-ancestry polygenic prediction method that uses a Bayesian framework with a continuous shrinkage prior to estimate SNP effect sizes from GWAS summary statistics across different populations. LD reference panels for EUR and EAS, provided by the PRS-CSx software, were constructed from the 1000 Genomes Project (1KG) samples. We used the default setting with the gamma-gamma prior hyperparameters (a and b) at 1 and 0.5, respectively. The shrinkage parameter $\phi$ was assessed at $10^{-6}$, $10^{-4}$, $10^{-2}$, and 1 within the tuning dataset to select the value with highest AUC. With the optimal $\phi$, we calculated the PRS using posterior effect-sizes from both EAS and EUR populations. Weights were then estimated to linearly combine the two PRSs on the tuning dataset. The final performance was then evaluated on an EAS independent validation dataset.

CT-SLEB: We used the recently developed CT-SLEB method to derive a PRS using EAS and EUR datasets. This method extends the standard CT method by designing a two-dimensional approach to select SNPs for EAS PRS construction. It incorporates an empirical Bayesian (EB) framework to model genetic correlations between EAS and EUR. Following this, a super-learning (SL) model is then applied to integrate multiple PRSs, each generated using distinct p-value thresholds and clumping parameters. Our implementation of CT-SLEB followed the default
setting with p-value thresholds $p_t = 5 \times 10^{-8}, 5 \times 10^{-7}, \ldots, 5 \times 10^{-1}$, or 1, and genetic distances $d = 50/r^2$ or $100/r^2$, where $r^2 = 0.01, 0.05, 0.1, 0.2, 0.5$ or 0.8. The clumping process for EAS was based on the LD reference derived from the FLCCA tuning dataset, and the EUR LD reference was based on the 1KG EUR samples.

**Relative and absolute risk calculation for PRSs**

To evaluate each PRS’s performance, we standardized the scores to a unit standard deviation, based on the control group in the EAS validation dataset. We then quantified the association between the standardized PRSs and LUAD risk using the OR and 95% CI, via logistic regression, adjusting for age and the first 10 principal components (PCs). We evaluated the predictive performance of the different PRS models through the adjusted AUC values, accounting for age and top 10 PCs, using the R package RISCA. We further evaluated the ORs of each PRS for risk of LUAD based on PRS percentiles, setting the middle quantile (40-60%) as the reference category.

For absolute risk estimates, we used the iCARE software to calculate the cumulative lifetime (age 30-80) and 10-year absolute risks of LUAD among never-smoking female controls in FLCCA (N=4,544). Absolute risks were derived by applying the Cox proportional hazard model with the top performing PRS (CT-SLEB) and age-specific lung cancer incidence and mortality rates in Taiwan.

**Projections of CT and LDpred2 PRS performance by sample size**

We used GENESIS package to estimate the sample sizes needed for PRS to reach various AUC levels in the EAS population. This method estimates the expected number of SNP discoveries and their explained heritability in future studies. Using the GWAS summary statistics from our EAS training dataset, and the provided LD scores for EAS populations from the 1KG dataset, we projected the AUC for CT PRS across various case-control ratios, from 1:1 to 1:10, and case
numbers ranging from 5,000 to 200,000. Given GENESIS’s specific design for CT PRS, we developed a extend its application to LDpred2 PRS projections. This involved modeling the relationship between effective sample sizes with the phenotypic variance ratio between LDpred2 and CT PRS (Supplementary Figure 1), elaborated in the Supplementary Note.

Results

Development and validation of the PRS

We applied several cutting-edge single- and multi-ancestry PRS methods (Methods, Supplementary Figure 2), evaluating their performance in terms of relative risk and AUC within the EAS validation dataset (Methods, Table 1, Supplementary Figure 2-3). Among single-ancestry PRS methods, the LDpred2 PRS, incorporating 942,591 SNPs, outperformed other methods, with an odds ratio (OR) per unit standard deviation (SD) of 1.62 (95% confidence interval (CI): 1.52, 1.73) alongside an adjusted AUC of 0.629 (95% CI: 0.618, 0.641). In contrast, the PRS designed exclusively for EUR populations underperformed in our EAS population, with an estimated adjusted AUC of 0.489 (95% CI: 0.477, 0.501), possibly attributed to the inclusion of smokers in the development of the existing EUR PRS.

Among multi-ancestry PRS methods, the weighted-PRS of EAS and EUR yielded an OR per unit SD of 1.62 (95% CI: 1.52, 1.72), with PRS-CSx and CT-SLEB showing even stronger association with ORs of 1.67 (95% CI: 1.57, 1.78) and 1.71 (95% CI: 1.61, 1.82), respectively. Among all the PRS methods, CT-SLEB had the best performance in the validation dataset with an AUC of 0.640 (95% CI: 0.629, 0.653). Setting individuals in the middle PRS quantile (40th to 60th percentile) as the reference category, those in the highest 5% of risk for the top performing PRS, CT-SLEB, had 4.17 (95% CI: 3.20, 5.47)-fold risk, whereas those in the lowest 5% had 0.33 (95% CI: 0.21, 0.51)-fold risk of developing LUAD (Figure 2).

Absolute risk of developing lung adenocarcinoma by CT-SLEB PRS
Using the iCARE package\textsuperscript{38}, we estimated the absolute risk of LUAD for never-smoking EAS women, utilizing the CT-SLEB PRS which demonstrated the highest AUC (Table 1). The cumulative lifetime risk of LUAD, assessed between age 30-80, varied markedly across PRS percentiles, ranging from a minimal 0.78% in the 5th percentile to a substantial 6.69% in the 95th percentile (Figure 3). Additionally, the 10-year absolute risk for LUAD in a never-smoking 50-year-old woman, a critical age for initiating recommended annual lung cancer screening as per the United States Preventive Services Taskforce (USPSTF), varied from 0.13% in the 5th percentile to 1.11% in the 95th percentile of the PRS distribution, emphasizing the PRS's potential in guiding lung cancer screening strategies.

\textit{Projections of polygenic risk score performance by sample size}

Using the GENESIS model, we projected the expected AUC of PRS-CT and LDpred2 PRS under varying GWAS sample sizes and case-control ratios for LUAD in never smoking EAS women (Figure 4). Our analysis, using the EAS training dataset, estimated approximately 1,772 (s.e.=1,641) susceptibility variants that are independently associated with LUAD\textsuperscript{27}. This high number of variants underscores the extensive polygenic nature of LUAD, implying relatively small effect sizes for individual SNPs. Based on our current data, the expected AUC for the LDpred2 PRS stands at 0.631 (95\% CI: 0.618, 0.641), aligning closely with the actual performance of the LDpred2 PRS in the validation dataset. As our projection increases in sample size to 65,000 cases with a one-to-one case-control ratio and a one-to-ten case-control ratio, the AUC is predicted to rise to 0.681 (95\% CI: 0.601, 0.726) and 0.704 (95\% CI: 0.629, 0.728), respectively. The upper boundary of our predictions, based solely on PRS, suggests an AUC potential of up to 0.731 (95\% CI: 0.640,0.786).

\textit{Explaining genetic variance through PRS across different sample sizes}
We also evaluated the genetic variance explained by the PRS under different sample sizes given the projected results (Supplementary Note, Table 2). The current LDpred2 PRS explained 26.6% of the genome-wide chip heritability, contributing to approximately 16.5% of the 1.84-fold familial relative risk for lung cancer among EAS never-smokers. Concurrently, genome-wide chip variants heritability explained 61.9% of the 1.84-fold familial relative risk associated with the disease. With an expanded sample size of 35,000 cases and 350,000 controls, the constructed PRS is projected to account for 57.9% of the genome-wide chip variants heritability, and 35.8% of the 1.84-fold familial relative risk for the disease.

Discussion

We developed and validated single- and multi-ancestry PRSs for LUAD in never-smoking EAS individuals using the largest GWAS dataset of never-smokers to date. The multi-ancestry PRS method, CT-SLEB, integrating summary data from EAS and EUR never-smokers, emerged as the best-performing PRS. It exhibited a dose-response relationship with LUAD risk and achieved higher AUC than all other evaluated PRSs. Further, our analysis demonstrates the potential of PRS in stratifying individuals' 10-year and lifetime risk of developing LUAD. Lastly, we projected the expected discriminatory accuracy of the PRS across a range of sample sizes and case-control ratios.

To date, GWAS and subsequent PRS models have largely centered on EUR populations⁴¹,⁴², rendering them less precise when applied to non-EUR populations and risking the exacerbation of health disparities⁴³,⁴⁴. Consistent with prior research, our study observed that a lung cancer PRS constructed using EUR data significantly underperformed when applied to EAS individuals¹⁵. However, by utilizing novel methods to integrate GWAS data from both EUR and EAS populations, we demonstrated that a multi-ancestry PRS enhances the precision of risk stratification for LUAD among EAS never-smokers.
While large-scale efforts to integrate PRS into clinical practice have focused on conditions with well-established PRS prediction performance\(^{16-18}\), such as Type 2 Diabetes, breast cancer, and cardiovascular diseases, lung cancer has been notably absent from these efforts, primarily due to the overwhelming focus on smoking as a risk factor. This has limited the exploration and application of PRS in lung cancer risk stratification, particularly among never-smokers. Our study addresses this gap by presenting a PRS that signifies a 3.92-fold increase in lung cancer risk for individuals in the top 20% risk quantile, a marked improvement over the 2.09-fold increase observed in the prior study by Wei et al.\(^{45}\) for a cohort of Chinese never-smoking women.

The CT-SLEB method’s improved performance can be attributed to its ability to utilize diverse genotyping arrays, thereby including population-specific variants. This aspect, along with the inclusion of EUR samples in the training set, broadens the genetic diversity and leverages larger sample sizes for improved model accuracy. Moreover, CT-SLEB excels in modeling genetic correlations across ancestries, efficiently using both shared and unique genetic markers for refined disease risk estimations across populations. Lastly, we set up our study with a robust three-sample split design, dividing datasets into training, tuning, and validation. This ensures that PRS efficacy is validated independently, effectively reducing the overfitting risk.

Current PRS studies predominantly focus on evaluating relative and absolute risks\(^{8,12,15,45,46}\), yet our research introduces a crucial forecast of sample sizes required for differing levels of PRS accuracy. Our projection analyses indicated that achieving an AUC of 0.70 would approach the maximized prediction potential of the PRS, given the estimated genetic variance explained by GWAS chip variants. To reach this level of accuracy, a future study would need to include 65,000 EAS LUAD cases with a 1:4 case-control ratio, or 55,000 cases with a 1:10 case-control ratio. Notably, accessing large biobanks with publicly available controls could reduce the number of required cases\(^{47}\). These projections offer a strategic framework for planning and designing future genetic studies on lung cancer, establishing clear benchmarks for PRS model performance, which is essential for the scientific community in advancing genetic risk prediction.
Our study has several limitations. First, it focuses on genetic susceptibility without considering additional questionnaire-based risk factors, aiming primarily to identify the most effective PRS model within genetic contexts. We plan to conduct integrative analyses that merge PRS with other risk indicators to refine LUAD risk predictions for never-smokers. Second, our projections are tailored to single-ancestry PRS models, not fully addressing the intricacies of multi-ancestry approaches, including diverse sample sizes and genetic correlations across populations. Third, our validation cohort, derived from the FLCCA and covering various EAS regions, relied on Taiwanese incidence data for absolute risk estimations. Consequently, our absolute risk findings may not extend universally to all EAS never-smokers, though our PRS performance assessment and sample size projections remain applicable to our population of interest. Lastly, the clinical application of our PRS findings, particularly in screening and risk counseling, is still in preliminary stages. Despite advancements in using PRS for smoking cessation trials, its implementation for never-smokers in clinical settings has been limited, presenting a significant area for future research.

In summary, our study evaluates various PRS models to capture the genetic predisposition to LUAD among never-smoking EAS individuals. It extends beyond risk prediction by estimating both 10-year and cumulative lifetime absolute risks, and by projecting the sample sizes required for future GWAS to refine the predictive power of PRSs in future GWAS. Additionally, we quantify the phenotypic variance captured by PRSs across different sample sizes. Future studies are crucial to further improve these PRS models, aiming to enhance genetic risk predictions while integrating a wider array of risk factors. Such efforts will develop more accurate and comprehensive risk models for LUAD in never-smoking individuals across diverse populations.
Acknowledgements

This work utilized the computational resources of the high-performance computation Biowulf cluster at National Institutes of Health, USA (http://hpc.nih.gov).

Female Lung Cancer Consortium in Asia (NCI): This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports, Culture and Technology of Japan, a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry Health, Labor and Welfare of Japan, by Health and Labor Sciences Research Grants for Research on Applying Health Technology from the Ministry of Health, Labor and Welfare of Japan, by the National Cancer Center Research and Development Fund, the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (grant No. 2011-0016106), a grant of the National Project for Personalized Genomic Medicine, Ministry for Health & Welfare, Republic of Korea (A111218-11-GM04), the Program for Changjiang Scholars and Innovative Research Team in University in China (IRT_14R40 to K.C.), the National Science & Technology Pillar Program (2011BAI09B00), MOE 111 Project (B13016), the National Natural Science Foundation of China (No. 30772531, and 81272618), Guangdong Provincial Key Laboratory of Lung Cancer Translational Medicine (No. 2012A061400006), Special Fund for Research in the Public Interest from the National Health and Family Planning Commission of PRC (No. 201402031), and the Ministry of Science and Technology, Taiwan (MOST 103-2325-B-400-023 & 104-2325-B-400-012). The Japan Lung Cancer Study (JLCS) was supported in part by the Practical Research for Innovative Cancer Control from Japan Agency for Medical Research and Development (15ck0106096h0002) and the Management Expenses Grants from the Government to the National Cancer Center (26-A-1) for Biobank. BioBank Japan was supported by the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government. The Japan Public Health Center-based prospective Study (the JPHC Study) was supported by the National Cancer Center Research and Development Fund (23-A- 31[toku], 26-A-2, 29-A-4, 2020-J-4, and 2023-J-4) (since 2011) and a Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan (from 1989 to 2010). The Taiwan GELAC Study (Genetic Epidemiological Study for Lung AdenoCarcinoma) was supported by grants from the National Research Program on Genomic...
Medicine in Taiwan (DOH99-TD-G-111-028), the National Research Program for Biopharmaceuticals in Taiwan (MOHW 103-TDUPB-211-144003, MOST 103-2325-B-400-023) and the Bioinformatics Core Facility for Translational Medicine and Biotechnology Development (MOST 104-2319-B-400-002). This work was also supported by the Jinan Science Research Project Foundation (201102051), the National Key Scientific and Technological Project (2011ZX09307-001-04), the National Natural Science Foundation of China (No.81272293), the State Key Program of National Natural Science of China (81230067), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2014R1A2A2A05003665), Sookmyung Women’s University Research Grants, Korea (1-1603-2048), Agency for Science, Technology and Research (A*STAR), Singapore and the US National Institute of Health Grant (1U19CA148127-01). The overall GWAS project was supported by the intramural program of the US National Institutes of Health/National Cancer Institute. The following is a list of grants by study center: SKLCS (Y.T.K.)—National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (2011-0016106). (J.C.) – This work was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health &Welfare, Republic of Korea (grant no. 0720550-2). (J.S.S) – grant number is A010250. WLCS (T.W.)—National Key Basic Research and Development Program (2011CB503800). SLCS (B.Z.)—National Nature Science Foundation of China (81102194). Liaoning Provincial Department of Education (LS2010168). China Medical Board (00726). GDS (Y.L.W.)—Foundation of Guangdong Science and Technology Department (2006B060101010, 2007A032000002, 2011A030400010). Guangzhou Science and Information Technology Bureau (2011Y2-00014). Chinese Lung Cancer Research Foundation, National Natural Science Foundation of China (81101549). Natural Science Foundation of Guangdong Province (S2011010000792). TLCS (K.C., B.Q)—Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT), China (IRT1076). Tianjin Cancer Institute and Hospital. National Foundation for Cancer Research (US). FLCS (J.C.W., D.R., L.J.)—Ministry of Health (201002007). Ministry of Science and Technology (2011BAI09B00). National S&T Major Special Project (2011ZX09102-010-01). China National High-Tech Research and Development Program (2012AA02A517, 2012AA02A518). National Science Foundation of China (30890034). National Basic Research Program (2012CB944600).
Scientific and Technological Support Plans from Jiangsu Province (BE2010715). NLCS (H.S.)—China National High-Tech Research and Development Program Grant (2009AA022705). Priority Academic Program Development of Jiangsu Higher Education Institution. National Key Basic Research Program Grant (2011CB503805). GEL-S (A.S.)—National Medical Research Council Singapore grant (NMRC/0897/2004, NMRC/1075/2006). (J.Liu)—Agency for Science, Technology and Research (A*STAR) of Singapore. GELAC (C.A.H.)—National Research Program on Genomic Medicine in Taiwan (DOH98-TDG-111-015). National Research Program for Biopharmaceuticals in Taiwan (DOH 100- TD-PB-111-TM013). National Science Council, Taiwan (NSC 100-2319-B-400-001). YLCS (Q.L.)—Supported by the intramural pro- gram of U.S. National Institutes of Health, National Cancer Institute. SWHS (W.Z., W-H.C., N.R.)—The work was supported by a grant from the National Institutes of Health (R37 CA70867, UM1 CA182910) and the National Cancer Institute intramural research program, including NCI Intramural Research Program contract (N02 CP1101066). JLCS (K.M., T.K.)—Grants-in-Aid from the Ministry of Health, Labor, and Welfare for Research on Applying Health Technology and for the 3rd-term Comprehensive 10-year Strategy for Cancer Control; by the National Cancer Center Research and Development Fund; by Grant-in-Aid for Scientific Research on Priority Areas and on Innovative Area from the Ministry of Education, Science, Sports, Culture and — Technology of Japan. (W.P.)—NCI R01-CA121210. HKS (J.W.)—General Research Fund of Research Grant Council, Hong Kong (781511M). The Environment and Genetics in Lung Cancer Etiology (EAGLE), Prostate, Lung, Colon, Ovary Screening Trial (PLCO), and Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) studies were supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute (NCI), Division of Cancer Epidemiology and Genetics. ATBC was also supported by U.S. Public Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-RC-37004) from the NCI. PLCO was also supported by individual contracts from the NCI to the University of Colorado Denver (N01-CN-25514), Georgetown University (N01-CN-25522), the Pacific Health Research Institute (N01-CN-25515), the Henry Ford Health System (N01-CN-25512), the University of Minnesota, (N01-CN-25513), Washington University (N01-CN-25516), the University of Pittsburgh (N01-CN-25511), the University of Utah (N01-CN-25524), the Marshfield Clinic Research Foundation (N01-CN-25518), the University of Alabama at
Birmingham (NO1-CN-75022), Westat, Inc. (NO1-CN-25476), and the University of California, Los Angeles (NO1-CN-25404). The Carotene and Retinol Efficacy Trial (CARET) is funded by the National Cancer Institute, National Institutes of Health through grants U01-CA063673, UM1-CA167462, and U01-CA167462. The Cancer Prevention Study-II (CPS-II) Nutrition Cohort was supported by the American Cancer Society. The NIH Genes, Environment and Health Initiative (GEI) partly funded DNA extraction and statistical analyses (HG-06-033-NCI-01 and RO1HL091172-01), genotyping at the Johns Hopkins University Center for Inherited Disease Research. This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C4002236).

Female Lung Cancer Consortium in Asia (Tianjin): Tianjin Science and Technology Committee Foundation, 18YFZCSY00520.


The GWAS of lung cancer in European never smokers was supported by NIH R01 CA149462 (OYG).

OncoArray study in Europeans: The OncoArray data and analysis from INTEGRAL-ILCCO were supported by NIH U19 CA203654, and U19 CA148127. The data harmonization for ILCCO was supported by Canadian Institute for Health Research (CIHR) Canada Research Chair to R.J.H, and CIHR FDN 167273).
European never-smoking lung cancer study: CIA is a Research Scholar of the Cancer Prevention Institute of Texas (CPRIT) and supported by CPRIT grant RR170048.

Taiwan eQTL study: This study was supported by the Ministry of Health and Welfare grants DOH97-TD-G-111-028 (ISC), DOH98-TD-G-111-017 (ISC), DOH99-TD-G-111-014 (ISC); DOH97-TD-G-111-026 (CAH), DOH98-TD-G-111-015 (CAH), DOH99-TD-G-111-028 (CAH); National Health Research Institutes grants NHRI-PH-110-GP-01, NHRI-PH-110-GP-03; and the Ministry of Science and Technology grants MOST108-2314-B-400-038(CAH), MOST109-2740-B-400-002(CAH).

B.B., K.W., J.C., J.S. N.R., Q.L. and H.Z. are supported by NIH intramural Research Program. N.C. is supported by NIH grant 1R01HG010480. P.Y. is supported by Mayo Clinic Foundation Research Funds, NIH-CA77118 and CA80127. G.L. is supported is supported by the Alan Brown Chair and Lusi Wong Fund of the Princess Margaret Cancer Foundation. D.C.C is supported by U01CA209414. O.Y.G. is supported by NIH R01 CA231141.

**Author contributions**

O.Y.G., I.P.G., X.W., PY. conducted epidemiology studies and contributed samples to GWAS
and/or conducted initial genotyping. All authors reviewed and approved the final review of the
manuscript.

**Competing interests**

The authors declare no competing interests.
### Table 1. Prediction performance of different methods for generating polygenic risk scores for lung cancer in never-smoking East Asian populations.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Number of SNPs</th>
<th>OR (95% CI)(^1)</th>
<th>AUC (95% CI)(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single-ancestry method(^3)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRS-25(^4)</td>
<td>25 (24 available)</td>
<td>1.53 (1.44, 1.63)</td>
<td>0.621 (0.612, 0.637)</td>
</tr>
<tr>
<td>PRS-CT</td>
<td>8</td>
<td>1.37 (1.29, 1.46)</td>
<td>0.591 (0.585, 0.609)</td>
</tr>
<tr>
<td>LDpred2 PRS</td>
<td>942,591</td>
<td>1.62 (1.52, 1.73)</td>
<td>0.629 (0.618, 0.641)</td>
</tr>
<tr>
<td>PRS-EUR 128(^5)</td>
<td>128 (109 available)</td>
<td>0.97 (0.91, 1.03)</td>
<td>0.489 (0.477, 0.501)</td>
</tr>
<tr>
<td><strong>Multi-ancestry method(^6)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDpred2 PRS + PRS-EUR 128</td>
<td>942,700</td>
<td>1.62 (1.52, 1.72)</td>
<td>0.629 (0.617, 0.640)</td>
</tr>
<tr>
<td>PRS-Csx</td>
<td>969,720</td>
<td>1.67 (1.57, 1.78)</td>
<td>0.637 (0.625, 0.647)</td>
</tr>
<tr>
<td>CT-SLEB</td>
<td>2,127,229</td>
<td>1.71 (1.61, 1.82)</td>
<td>0.640 (0.629, 0.653)</td>
</tr>
</tbody>
</table>

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; AUC, area under the curve; PC, principle component; CT, clumping and thresholding.

\(^1\) ORs are provided per one unit standard deviation of PRS.

\(^2\) Adjusted for age (continuous) and first 10 principal components.

\(^3\) Sample size for single-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian (EAS) ancestry for training, 2,219 cases and 2,272 controls of EAS ancestry for tuning and 2,219 cases and 2,272 controls of EAS ancestry for validation.

\(^4\) SNPs that reached genome-wide significance in Shi et al., 2023.

\(^5\) SNPs obtained from genome-wide association study conducted in individuals with European ancestry (Hung et al., 2021, Cancer Res.).

\(^6\) Sample size for multi-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian (EAS) ancestry, as well as 2,058 cases and 5,575 controls of European ancestry for training, 2,219 cases and 2,272 controls of EAS ancestry for validation.
Table 2. Genetic variance in East Asian lung cancer among never smokers explained by LDpred2 PRS

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Genetic variance of PRS²</th>
<th>Proportion of all-GWAS variants genetic variance explained by PRS³</th>
<th>Proportion of Familial risk explained by PRS⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>0.201</td>
<td>26.6%</td>
<td>16.5%</td>
</tr>
<tr>
<td>35,000 cases</td>
<td>0.437</td>
<td>57.9%</td>
<td>35.8%</td>
</tr>
<tr>
<td>55,000 cases</td>
<td>0.540</td>
<td>71.5%</td>
<td>44.3%</td>
</tr>
</tbody>
</table>

¹ Sample Size: The current sample comprises 3,564 cases and 16,238 controls. Future projections assume a 1:10 case-control ratio for sample sizes of 35,000 and 55,000 cases. Genetic variance projections for the LDpred2 PRS are based on the GENESIS method, originally designed for the CT method (refer to Nat. Genet. 50, 1318-1326 (2018)), extended to include LDpred2 by modeling the variance ratio between LDpred2 and CT (Supplementary Note).

² Genetic Variance of PRS: This corresponds to the heritability on the frailty scale, assuming a polygenic log-additive model underpins this relationship. It quantifies the proportion of the phenotype variation that can be attributed to genetic factors in the context of PRS.

³ Proportion of Genetic Variance from All-GWAS Variants Explained by PRS: This represents the variance of all genome-wide imputable variants as established through LD-score regression (refer to Nat. Genet. 47, 291-5 (2015) and Nat. Genet. 47, 1236-41 (2015)). On the frailty scale, the genetic variance of all GWAS variants is calculated as \( \sigma^2_{GWAS} = Var(\sum_{m=1}^{M} \beta_m G_m) \), where \( G_m \) is the standardized genotype for the mth SNP, \( \beta_m \) is the true log odds ratio for the mth SNP and M is the total number of causal SNPs among the GWAS variants. For East Asian never smokers, the estimated genetic variance of all GWAS variants is 0.755.

⁴ Proportion of Familial Risk Explained by PRS: This calculates the familial risk in terms of genetic variance using the formula \( \lambda_s^2 = \exp(\sigma^2) \), where \( \lambda_s \) is the familial risk when a first-order sibling has the disease, and \( \sigma^2 \) is the genetic variance on frailty-scale. Further details of this calculation can be found in Nat. Genet. 31, 33-36 (2002). For lung cancer in East Asian never-smokers, the familial risk is a 1.84-fold increase, and the genetic variance of all GWAS variants, as estimated through LD-score regression, explains 61.9% of this increased familial relative risk.
Figure 1. Overview of data structure, polygenic risk score (PRS) development, validation and application. Summary statistics from East Asian (EAS) genome-wide association studies (GWAS) were used to develop single-ancestry PRS using methods such as a simple PRS constructed using 25 SNPs that have previously reached genome-wide significance (i.e., P<5x10^{-8}) (PRS-25), a PRS using the clumping and thresholding (CT) method (PRS-CT) incorporating 8 SNPs, and a PRS using a genome-wide Bayesian-based approach, LDpred2 (LDpred2 PRS) incorporating close to a
million SNPs. For the multi-ancestry PRS development, we also used summary statistics from European (EUR) GWAS, applying the PRS-CSx method that leveraged genome-wide association summary statistics for close to a million SNPs with a Bayesian continuous shrinkage prior to model SNP effect sizes across populations, as well as CT-SLEB method, which enhances the standard CT methods with a two-dimensional approach to select SNPS for EAS PRS construction by incorporating over 2 million SNPs. Tuning and validation of each PRS was conducted in an independent EAS individual-level data. Relative risk per PRS quantile was calculated as an odds ratio (OR) with the middle quantile (40th to 60th percentile) set as the reference, and the area under the receiver operating curve (AUC) was estimated for each PRS. CT-SLEB PRS was used to estimate 10-year and lifetime cumulative absolute risk, and PRS-CT and PRS-LDpred2 were used for sample size projection.
Figure 2. Relative risk estimated for quantiles of each polygenic risk score (PRS) and lung adenocarcinoma in the validation dataset of women with East Asian ancestry, treating the 40th to 60th percentile as the references. Odds ratios of PRS per standard deviation (SD) and 95% confidence intervals are shown for the single-ancestry 25 SNP polygenic risk score, PRS-25 (A), Clumping and thresholding method, PRS-CT (B), Bayesian-based genome-wide approach, LDpred2 PRS (C), multi-ancestry approach, CT-SLEB (D).
Figure 3. **Lifetime cumulative and 10-year absolute risk of developing lung cancer.** Lifetime Cumulative (A) 10-year (B) absolute risk of developing lung adenocarcinoma in never-smoking women in East Asia by percentiles of the CT-SLEB polygenic risk score (PRS). Absolute risks were calculated based on Taiwan’s age specific incidence and mortality data and using the PRS relative risks estimated as described in the material and methods section.
Figure 4. The expected area under the receiver operating characteristic curve (AUC) of a polygenic risk score (PRS) built based on a GWAS of specified sample sizes for lung adenocarcinoma in never-smoking East Asian women. Projected AUC values for PRS-CT and
LDpred2 PRS with a 1:1 and a 10:1 case to control ratio (A) and AUC values for LDpred2 PRS for 1:1, 2:1, 4:1, 5:1 and 10:1 case to control ratio (B).
References


**Data Availability**

All data supporting the findings described in this paper are available in the paper and in the Supplementary Information and from the corresponding author or as otherwise indicated upon request. The individual genotype data for the FLCCA data are in dbGaP phs000716.v1.p1 (Genome-Wide Association Study of Lung Cancer Susceptibility in Never-Smoking Women in Asia). For the NCC and ACC studies, please contact Kouya Shiraishi at kshirais@ncc.go.jp or Takashi Kohno at tkkohno@ncc.go.jp for summary statistics. The GWAS data for the European populations contributing to this study are available at dbGap under accession phs000877.v1.p1 (Transdisciplinary Research Into Cancer of the Lung (TRICL), https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000876.v2.p1), phs001273.v3.p2 (Oncoarray Consortium, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-
To gain access to all data in dbGaP cited in this paper, please apply for dbGaP Authorized Access. GWAS data for the European populations contributing to this study are available under accession phs000877.v1.p1 (ILLCO), phs001273.v1.p1 (OncoArray).

Code Availability


Most of our statistical analyses were performed using the following R packages: ggplot2 v.3.3.3, dplyr v.1.0.4, data.table v.1.13.6, iCARE v.1.30.0 https://github.com/KevinWFred/PRS_EASLC

1Department of Clinical Epidemiology and Center of Evidence Based Medicine, The First Hospital of China Medical University, Shenyang, China, 2Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China, 3State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China, 4Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA, 5Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore, 6Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore, 7Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan, 8Department of Applied Mathematics, Chung Yuan Christian University, Chong-Li, Taiwan, 9Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center and Vanderbilt-Ingram Cancer Center, Nashville, TN, USA, 10Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea, 11Guangdong Lung Cancer Institute, Guangdong Provincial Key Laboratory of Translational Medicine in Lung Cancer, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China, 12Department of Pathology, Queen Mary Hospital, Hong Kong, Hong Kong, 13Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA, 14Department of Biostatistics, Bioinformatics & Biomathematics, Georgetown University Medical Center,
Washington, DC, USA, Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA, Saville Cancer Screening and Prevention Center, Inova Schar Cancer Institute, Inova Health System, Fairfax, VA, USA, Department of Environmental Health, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea, Epidemiology and Community Health Branch, National Heart Lung and Blood Institute, Bethesda, MD, USA, Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea, Shanghai Pulmonary Hospital, Shanghai, China, Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China, Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan, Lung and Esophageal Cancer Clinic, Chonnam National University Hwasun Hospital, Hwasun eup, Republic of Korea, Department of Internal Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea, Department of Etiology & Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, Department of Medicine, School of Clinical Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong, Hong Kong, Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, Center for Antibody and Vaccine Therapy, Research Hospital, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, Department of Medical Oncology and Cancer Center, and Center for Advanced Medicine against Cancer, Shiga University of Medical Science, Shiga, Japan, Department of Statistics & Research Institute of Natural Sciences, Sookmyung Women's University, Seoul, 04310, Korea, Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, Department of Thoracic Surgery, Tokyo Medical and Dental University, Tokyo, Japan, Department of Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, USA, Department of Medical Oncology, Japanese Red Cross Medical Center, Tokyo, Japan, Department of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan, Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, Yokohama, Japan, Department of Thoracic Surgery, Kanagawa Cancer Center, Yokohama, Japan, Department of Thoracic Oncology, National Cancer Center Hospital East, Kashiwa, Japan, Department of Thoracic Surgery, National Cancer Center Hospital East, Kashiwa, Japan, Division of Medical AI Research and Development, National Cancer Center Research Institute, Tokyo, Japan, Department of Cellular and Organ Pathology, Graduate School of Medicine, Akita University, Akita, Japan, Department of Thoracic Surgery, Graduate School of Medicine, Akita University, Akita, Japan, Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, Japan, Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan, Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Science, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, Department of Molecular Biology, Institute for Advanced Medical Sciences, Nippon Medical School, Tokyo, Japan, Department of Surgery, Division of General Thoracic Surgery, Shinshu University School of Medicine Asahi, Nagano, Japan, Department of Chest Surgery, Fukushima Medical University School of Medicine, Fukushima, Japan, Department of Gastrointestinal Tract Surgery, Fukushima Medical University School of
Medicine, Fukushima, Japan, Department of Integrative center of General Surgery, Gunma University Hospital, Gunma, Japan, Institute of Occupational Medicine and Ministry of Education Key Lab for Environment and Health, School of Public Health, Huazhong University of Science and Technology, Wuhan, China, China National Environmental Monitoring Center, Beijing, China, Department of Internal Medicine, Division of Oncology/Hematology, College of Medicine, Korea University Anam Hospital, Seoul, Republic of Korea, Department of Clinical Oncology, The University of Hong Kong, Queen Mary Hospital, Hong Kong, Hong Kong, School of Medicine and Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan, Department of Internal Medicine, Division of Pulmonary Medicine, Chung Shan Medical University Hospital, Taichung, Taiwan, Institute of Biomedical Sciences, National Chung Hsing University, Taichung, Taiwan, Department of Internal Medicine, Division of Chest Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, Department of Respiratory Therapy, Chang Gung University, Taoyuan, Taiwan, Department of Pulmonary and Critical Care, Xiamen Chang Gung Hospital, Xiamen, China, Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan, Department of Internal Medicine, E-Da Cancer Hospital, I-Shou University and Kaohsiung Medical University, Kaohsiung, Taiwan, Department of Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan, Department of Chest Medicine, Taipei Veterans General Hospital, and school of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, Lung Cancer Center, Kyungpook National University Medical Center, Daegu, Republic of Korea, Jeonnam Regional Cancer Center, Chonnam National University, Hwasun, Republic of Korea, Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China, Department of Etiology & Carcinogenesis and State Key Laboratory of Molecular Oncology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, Department of Human Genetics, Genome Institute of Singapore, Singapore, Singapore, School of Life Sciences, Anhui Medical University, Hefei, China, Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA, Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK, Department of Medicine, Section of Epidemiology and Population Science, Institute for Clinical and Translational Research, Houston, TX, USA, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, USA, Department of Integrative Oncology, BC Cancer, Vancouver, BC, Canada, IUOPA, University of Oviedo and CIBERESP, Spain, Program in Epidemiology, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark, International Agency for Research on Cancer (IARC/WHO), Lyon, France, German Cancer Research Center (DKFZ), Heidelberg, Germany, Translational Lung Research Center Heidelberg (TLRC-H), Member of the German Center for Lung Research (DZL), Heidelberg, Germany, University of Salzburg and Cancer Cluster Salzburg, Salzburg, Austria, University Medical Center Goettingen, Goettingen, Germany, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany, Helmholtz Center Munich, Institute of Epidemiology, Munich, Germany, Institute of Medical Statistics and Epidemiology, Technical University Munich, Munich, Germany, Harvard TH Chan School of Public Health, Boston,
Massachusetts, USA, 91Carmel Medical Center, Israel, 92Markey Cancer Center, Lexington, KY, USA, 93Liverpool University, Liverpool, UK, 94The University of Texas MD Anderson Cancer Center, Houston, Texas, USA, 95Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA, 96Princess Margaret Cancer Center, Toronto, ON, Canada, 97Norris Cotton Cancer Center, Lebanon, NH, USA, 98Radboud University Medical Center, Nijmegen, Netherlands, 99National Institute of Occupational Health, Oslo, Norway, 100Department of Medical Biosciences, Umeå University, Umeå, Sweden, 101Department of Radiation Sciences, Umeå University, Umeå, Sweden, 102University of Sheffield, Sheffield, UK, 103UPMC Hillman Cancer Center and Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA, 104Department of Pharmaceutical Sciences, College of Pharmacy and Pharmaceutical Sciences, Washington State University, Spokane, WA, USA, 105Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA, 106Department of Thoracic Surgery, Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA, 107Cancer Research Center, Kyungpook National University Medical Center, Daegu, Republic of Korea, 108National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan, 109Department of Thoracic and Cardiovascular Surgery, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea, 110Department of Occupational and Environmental Health and Ministry of Education Key Lab for Environment Health and School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, 111Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China, 112Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc, Rockville, MD, USA, 113Department of Epidemiology, School of Public Health, China Medical University, Shenyang, China, 114Division of Applied Oral Sciences & Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China, 115State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, China, 116State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, China, 117Department of Oncology, Cancer Center, Union Hospital, Huazhong University of Science and Technology, Wuhan, China, 118The Ph.D. Program for Translational Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan, 119School of Public Health, Li Ka Shing (LKS) Faculty of Medicine, The University of Hong Kong, Hong Kong, China, 120Key Laboratory of Cancer Etiology and Intervention, University of Liaoning Province, Shenyang, China, 121Department of Internal Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan, 122Gleneagles Hong Kong Hospital, Hong Kong, China, 123Department of Internal Medicine, Division of Medical Oncology, College of Medicine, Korea University Guro Hospital, Seoul, Republic of Korea, 124Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea, 125Department of Oncology, Wuhan Iron and Steel (Group) Corporation Staff-Worker Hospital, Wuhan, China, 126Division of Pulmonary Oncology and Interventional Bronchoscopy, Department of Thoracic Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan, 127Division of Thoracic Surgery, Department of Surgery, Chung Shan Medical University Hospital, Taichung, Taiwan, 128Division of Environmental Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands, 129Department of Epidemiology and Biostatistics, University of
California, San Francisco, San Francisco, CA, USA, 130Department of Pathology, Queen Elizabeth Hospital, Hong Kong, China, 131Qujing Center for Diseases Control and Prevention, Qujing, China, 132Department of Oncology, Shandong Cancer Hospital and Institute, Shandong Academy of Medical Sciences, Jinan, China, 133Division of Cohort Research, National Cancer Center Institute for Cancer Control, Tokyo, Japan, 134Division of Epidemiology, National Cancer Center Institute for Cancer Control, Tokyo, Japan, 135Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China, 136Center for Global Health, National Cancer Institute, Bethesda, MD, USA, 137Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan, 138Genomic Research Center, Academia Sinica, Taipei, Taiwan, 139Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan, 140Laboratory Services, American Cancer Society, Georgia, USA