

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

3  
4 Batel Blechter<sup>†,1</sup>, Xiaoyu Wang<sup>†,1,2</sup>, Jianxin Shi<sup>†,1</sup>, Kouya Shiraishi<sup>†,3</sup>, Jiyeon Choi<sup>†,1</sup>, Keitaro  
5 Matsuo<sup>†,4</sup>, Tzu-Yu Chen<sup>†,5,6</sup>, Juncheng Dai<sup>†,7,8</sup>, Rayjean J Hung<sup>†,9</sup>, Kexin Chen<sup>†,10</sup>, Xiao-Ou Shu  
6<sup>†,11</sup>, Young Tae Kim<sup>12,13</sup>, Parichoy Pal Choudhury<sup>14</sup>, Jacob Williams<sup>1</sup>, Maria Teresa Landi<sup>1</sup>, Dongxin  
7 Lin<sup>15</sup>, Wei Zheng<sup>11</sup>, Zhihua Yin<sup>16</sup>, Bao Song<sup>17</sup>, I-Shou Chang<sup>18</sup>, Yun-Chul Hong<sup>19</sup>, Never Smoker  
8 Lung Cancer Working Group, Nilanjan Chatterjee<sup>‡,20,21</sup>, Olga Y Gorlova<sup>‡,22,23</sup>, Christopher I Amos  
9<sup>‡,22,23</sup>, Hongbing Shen<sup>‡,7,8</sup>, Chao Agnes Hsiung<sup>‡,5</sup>, Stephen J Chanock<sup>‡,1</sup>, Nathaniel Rothman<sup>‡,1</sup>,  
10 Takashi Kohno<sup>‡,3</sup>, Qing Lan<sup>‡,1</sup>, Haoyu Zhang<sup>‡,1</sup>

11  
12 <sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA  
13 <sup>2</sup>Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos  
14 Biomedical Research Inc, Rockville, MD, USA  
15 <sup>3</sup>Division of Genome Biology, National Cancer Research Institute, Tokyo, Japan  
16 <sup>4</sup>Division of Molecular Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan  
17 <sup>5</sup>Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan  
18 <sup>6</sup>Department of Applied Mathematics, Tunghai University, Taichung, Taiwan  
19 <sup>7</sup>Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China  
20 <sup>8</sup>Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for  
21 Cancer Medicine, Nanjing Medical University, Nanjing, China  
22 <sup>9</sup>Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai  
23 Health, Toronto, Canada  
24 <sup>10</sup>Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key  
25 Laboratory of Molecular Cancer Epidemiology of Tianjin, Tianjin Medical University Cancer Institute and  
26 Hospital, Tianjin Medical University, Tianjin, China  
27 <sup>11</sup>Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center and  
28 Vanderbilt-Ingram Cancer Center, Nashville, TN, USA  
29 <sup>12</sup>Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea  
30 <sup>13</sup>Cancer Research Institute, Seoul National University College of Medicine, Seoul National University  
31 Hospital, Seoul, Republic of Korea

32 <sup>14</sup>American Cancer Society, Atlanta, GA, USA

33 <sup>15</sup>Department of Etiology & Carcinogenesis and State Key Laboratory of Molecular Oncology, Cancer  
34 Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing,  
35 China

36 <sup>16</sup>Department of Epidemiology, School of Public Health, China Medical University, Shenyang, China

37 <sup>17</sup>Department of Oncology, Shandong Cancer Hospital and Institute, Shandong Academy of Medical  
38 Sciences, Jinan, China

39 <sup>18</sup>National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan

40 <sup>19</sup>Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of  
41 Korea

42 <sup>20</sup>Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

43 <sup>21</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

44 <sup>22</sup>Department of Medicine, Section of Epidemiology and Population Science, Institute for Clinical and  
45 Translational Research, Houston, TX, USA

46 <sup>23</sup>Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA

47

48

49

50

51

52 † These authors contributed equally to this work

53 ‡ These authors jointly directed this work

54

55 Correspondence to: Batel Blechter ([batel.blechter@nih.gov](mailto:batel.blechter@nih.gov)), Qing Lan ([qingl@mail.nih.gov](mailto:qingl@mail.nih.gov)),  
56 and Haoyu Zhang ([haoyu.zhang2@nih.gov](mailto:haoyu.zhang2@nih.gov))

57

58

59

60

61

62 **Abstract**

63

64 **Background**

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

70

71 **Methods**

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

80

81 **Findings**

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

89

90 **Interpretations**

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

94

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

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

## 112 Introduction

113  
114 Lung cancer is a major global health challenge, responsible for about 20% of all cancer deaths in  
115 2020<sup>1</sup>. While smoking is the primary etiologic factor, around 25% of lung cancer cases occur in  
116 never-smokers, with significant geographical variations<sup>2</sup>. Notably, in East Asia (EAS), never-  
117 smoking women exhibit high incidences of lung adenocarcinoma (LUAD), the predominant  
118 histologic subtype<sup>3</sup>.

119  
120 Beyond smoking, risk factors for LUAD include environmental and occupational exposures,  
121 lifestyle, family history, and genetic susceptibility<sup>4-7</sup>. Genome-wide association studies (GWAS)  
122 for never-smoking lung cancer across EAS and European (EUR) populations have identified 28  
123 susceptibility variants at 25 independent loci<sup>8-13</sup>. These GWAS findings enhance our  
124 understanding of LUAD's genetic architecture, yet the translation of these findings into clinical  
125 practice requires further investigation.

126  
127 Polygenic risk scores (PRS) aggregate the effects of individual single nucleotide polymorphisms  
128 (SNPs) to estimate genetic risk for diseases, playing a crucial role in cancer prevention and  
129 screening. Previous efforts primarily utilized EUR populations for lung cancer PRS  
130 development<sup>14,15</sup> to better identify high-risk individuals. Nonetheless such PRSs often fall short  
131 for EAS never-smokers, reflecting a bias in screening guidelines and the focus of studies on  
132 populations with EUR ancestry. Despite ongoing efforts to integrate PRS into clinical practices  
133 for complex diseases, as seen in initiatives like the electronic Medical Records and Genomics  
134 (eMERGE) network<sup>16</sup>, Veterans Affairs Genomic Medicine at Veteran Affairs (GenoVA) study<sup>17</sup>,  
135 and the Women Informed to Screen Depending On Measures of risk (WISDOM) study<sup>18</sup>, lung  
136 cancer has been excluded. The predominant global burden of lung cancer, driven primarily by  
137 smoking, overlooks the significant impact and potential benefits of PRS in never-smoking  
138 populations.

139

140 The recent expansion of GWAS across diverse populations, combined with advancements in PRS  
141 methodology, underscores a pivotal shift towards enhancing disease outcome prediction  
142 beyond EUR populations<sup>19</sup>. Notably, the development of multi-ancestry PRS methods marks a  
143 considerable advancement, leveraging data from various ancestral backgrounds to enrich the  
144 predictive accuracy and robustness of PRSs<sup>20-26</sup>. By jointly modeling genetic data from multiple  
145 populations, these approaches improve the predictive power of PRSs, thereby enhancing  
146 disease outcome predictions in non-EUR populations.

147

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

154

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

164

165

166 **Methods**

167

168 *Study design and data sources*

169

170 Training data compilation: study population and genotyping

171 The studies, genotyping protocols, and quality control for PRS construction are previously  
172 described in detail<sup>27</sup>. Briefly, EAS single-ancestry PRSs for LUAD were constructed using  
173 summary data from 3,564 never-smoking LUAD cases and 16,238 never-smoking controls of EAS  
174 ancestry from the Nanjing Lung Cancer Study (NJLCS)<sup>9,28</sup>, National Cancer Center of JAPAN  
175 (NCC), and the Research Institute and Aichi Cancer Center (ACC). The NJLCS study combined  
176 data from several cities, genotyped by Affymetrix Genome-Wide Human SNP Array 6.0, as well  
177 as the Oncoarray GWAS<sup>28–30</sup>. The NCC study included lung cancer patients and control data from  
178 multiple Japanese, with genotyping conducted on Illumina HumanOmniExpress and  
179 HumanOmni1-Quad genotyping platforms. Similarly, the ACC study pooled data from the  
180 multiple Japanese medical institutions, and the Nagahama Study, with genotyping performed  
181 using Illumina 610k and Illumina660k platforms. To construct the multi-ancestry PRS, we further  
182 incorporated GWAS summary statistics from 2,058 never-smoking LUAD cases and 5,575 never-  
183 smoking controls of EUR ancestry, genotyped using Illumina Infinium OmniExpress-24 v1.2  
184 BeadChips and Illumina Human660W-Quad BeadChip<sup>13</sup>.

185

186 Tuning and validation: study population and genotyping

187 For PRS tuning and validation, we used 4,438 never-smoking LUAD cases and 4,544 never-  
188 smoking controls from the Female Lung Cancer Consortium in Asia (FLCCA)<sup>10,11</sup>, an international  
189 consortium composed of never-smoking EAS women from regions including Mainland China,  
190 Singapore, Taiwan, Hong Kong, South Korea and Japan<sup>10,11</sup>. All samples were genotyped using  
191 the Illumina 660W, 370K or 610Q microarrays. The FLCCA data were randomly and equally  
192 divided for tuning (2,219 cases, 2,272 controls) and validation (2,219 cases, 2,272 controls).  
193 Informed consent was obtained from all participants, and the recruitment and data collection

194 procedures were approved by the local ethics review committees of all participating research  
195 institutes.

196

197 *Construction of PRS*

198

199 Single-ancestry PRS methods

200 **PRS-25:** We constructed a simple PRS using the 25 independent SNPs (PRS-25) that reached  
201 genome-wide significance (i.e.,  $P < 5 \times 10^{-8}$ ) in the largest GWAS for LUAD in EAS population to  
202 date<sup>8</sup>. The PRS-25 was calculated by summing the risk alleles weighed by their effect sizes (i.e.,  
203 per allele log-odds ratio) obtained from a meta-analysis including only never-smokers.

204

205 **PRS-CT:** We generated a PRS using the CT method<sup>31</sup> using PLINK 1.90<sup>32</sup>. This involved the  
206 clumping of SNPs based on LD with a threshold of  $r^2 = 0.1$  within a 500kb window. The LD were  
207 estimated from half of the FLCCA dataset used for tuning dataset. We created nine SNP subsets  
208 by applying incremental P-value thresholds (i.e.,  $P < 5 \times 10^{-8}$ ,  $P < 5 \times 10^{-7}$ , ...,  $P < 5 \times 10^{-1}$ , 1), and  
209 calculated their respective PRSs using PLINK2<sup>32</sup> command “--score cols=+scoresums,-scoreavgs  
210 no-mean-imputation”. Using the tuning dataset, the AUC was calculated for each threshold to  
211 identify the most predictive P-value cutoff.

212

213 **LDpred2-PRS:** We applied the LDpred2 method, implemented in the R package bigsnpr<sup>33</sup>, using  
214 a Bayesian framework to estimate SNP effect sizes through a shrinkage estimator. This method  
215 leverages GWAS summary statistics, incorporating a prior for effect sizes, while also accounting  
216 for LD across SNPs. Our analysis was constrained to HapMap3 variants, and we calibrated the  
217 model using a range of hyperparameters: the proportion of causal SNPs was set across a 21-  
218 point logarithmic sequence from  $10^{-5}$  to 1, and the per-SNP heritability was set as a fraction  
219 (0.3, 0.7, 1 or 1.4) of the total heritability estimated by LD score regression<sup>34</sup> divided by the  
220 number of causal SNPs. Lastly, we used the “sparse” option to set weaker effects to zero. The  
221 optimal tuning parameters were selected based on the highest AUC achieved on the tuning  
222 dataset.



223

224 **PRS-EUR 128:** We evaluated the performance of a European-derived PRS (PRS-EUR 128)<sup>35</sup>,  
225 which consists of 128 variants obtained from a GWAS of predominately active smokers, on EAS  
226 populations to assess its cross-population applicability.

227

228 Multi-ancestry PRSs methods:

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

234

235 **PRS-CSx:** We applied PRS-CSx, a multi-ancestry polygenic prediction method that uses a  
236 Bayesian framework with a continuous shrinkage prior to estimate SNP effect sizes from GWAS  
237 summary statistics across different populations<sup>26</sup>. LD reference panels for EUR and EAS,  
238 provided by the PRS-CSx software, were constructed from the 1000 Genomes Project (1KG)  
239 samples<sup>36</sup>. We used the default setting with the gamma-gamma prior hyperparameters ( $\alpha$  and  
240  $\beta$ ) at 1 and 0.5, respectively. The shrinkage parameter  $\phi$  was assessed at  $10^{-6}$ ,  $10^{-4}$ ,  $10^{-2}$ , and 1  
241 within the tuning dataset to select the value with highest AUC. With the optimal  $\phi$ , we  
242 calculated the PRS using posterior effect-sizes from both EAS and EUR populations. Weights  
243 were then estimated to linearly combine the two PRSs on the tuning dataset. The final  
244 performance was then evaluated on an EAS independent validation dataset.

245

246 **CT-SLEB:** We used the recently developed CT-SLEB method to derive a PRS using EAS and EUR  
247 datasets<sup>24</sup>. This method extends the standard CT method by designing a two-dimensional  
248 approach to select SNPs for EAS PRS construction. It incorporates an empirical Bayesian (EB)  
249 framework to model genetic correlations between EAS and EUR. Following this, a super-learning  
250 (SL) model is then applied to integrate multiple PRSs, each generated using distinct p-value  
251 thresholds and clumping parameters<sup>37</sup>. Our implementation of CT-SLEB followed the default

252 setting with p-value thresholds  $p_t = 5 \times 10^{-8}, 5 \times 10^{-7}, \dots, 5 \times 10^{-1}$ , or 1, and genetic distances  
253  $d = 50/r^2$  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  
254 was based on the LD reference derived from the FLCCA tuning dataset, and the EUR LD  
255 reference was based on the 1KG EUR samples.

256

### 257 *Relative and absolute risk calculation for PRSs*

258

259 To evaluate each PRS's performance, we standardized the scores to a unit standard deviation,  
260 based on the control group in the EAS validation dataset. We then quantified the association  
261 between the standardized PRSs and LUAD risk using the OR and 95% CI, via logistic regression,  
262 adjusting for age and the first 10 principal components (PCs). We evaluated the predictive  
263 performance of the different PRS models through the adjusted AUC values, accounting for age  
264 and top 10 PCs, using the R package RISCA<sup>37</sup>. We further evaluated the ORs of each PRS for risk  
265 of LUAD based on PRS percentiles, setting the middle quantile (40-60%) as the reference  
266 category.

267

268 For absolute risk estimates, we used the iCARE software<sup>38</sup> to calculate the cumulative lifetime  
269 (age 30-80) and 10-year absolute risks of LUAD among never-smoking female controls in FLCCA  
270 (N=4,544). Absolute risks were derived by applying the Cox proportional hazard model with the  
271 top performing PRS (CT-SLEB) and age-specific lung cancer incidence and mortality rates in  
272 Taiwan<sup>39</sup>.

273

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

275

276 We used GENESIS package<sup>40</sup> to estimate the sample sizes needed for PRS to reach various AUC  
277 levels in the EAS population. This method estimates the expected number of SNP discoveries  
278 and their explained heritability in future studies. Using the GWAS summary statistics from our  
279 EAS training dataset, and the provided LD scores for EAS populations from the 1KG dataset, we  
280 projected the AUC for CT PRS across various case-control ratios, from 1:1 to 1:10, and case

281 numbers ranging from 5,000 to 200,000. Given GENESIS's specific design for CT PRS, we  
282 developed a extend its application to LDpred2 PRS projections. This involved modeling the  
283 relationship between effective sample sizes with the phenotypic variance ratio between  
284 LDpred2 and CT PRS (**Supplementary Figure 1**), elaborated in the **Supplementary Note**.

285

## 286 **Results**

287

### 288 *Development and validation of the PRS*

289

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

299

300 Among multi-ancestry PRS methods, the weighted-PRS of EAS and EUR yielded an OR per unit  
301 SD of 1.62 (95% CI: 1.52, 1.72), with PRS-CSx and CT-SLEB showing even stronger association  
302 with ORs of 1.67 (95% CI: 1.57, 1.78) and 1.71 (95% CI: 1.61, 1.82), respectively. Among all the  
303 PRS methods, CT-SLEB had the best performance in the validation dataset with an AUC of 0.640  
304 (95% CI: 0.629, 0.653). Setting individuals in the middle PRS quantile (40<sup>th</sup> to 60<sup>th</sup> percentile) as  
305 the reference category, those in the highest 5% of risk for the top performing PRS, CT-SLEB, had  
306 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)-  
307 fold risk of developing LUAD (**Figure 2**).

308

### 309 *Absolute risk of developing lung adenocarcinoma by CT-SLEB PRS*

310

311 Using the iCARE package<sup>38</sup>, we estimated the absolute risk of LUAD for never-smoking EAS  
312 women, utilizing the CT-SLEB PRS which demonstrated the highest AUC (**Table 1**). The  
313 cumulative lifetime risk of LUAD, assessed between age 30-80, varied markedly across PRS  
314 percentiles, ranging from a minimal 0.78% in the 5th percentile to a substantial 6.69% in the  
315 95th percentile (**Figure 3**). Additionally, the 10-year absolute risk for LUAD in a never-smoking  
316 50-year-old woman, a critical age for initiating recommended annual lung cancer screening as  
317 per the United States Preventive Services Taskforce (USPSTF), varied from 0.13% in the 5th  
318 percentile to 1.11% in the 95th percentile of the PRS distribution, emphasizing the PRS's  
319 potential in guiding lung cancer screening strategies.

320

#### 321 *Projections of polygenic risk score performance by sample size*

322

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

335

#### 336 *Explaining genetic variance through PRS across different sample sizes*

337

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

346

## 347 **Discussion**

348

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

357

358 To date, GWAS and subsequent PRS models have largely centered on EUR populations<sup>41,42</sup>,  
359 rendering them less precise when applied to non-EUR populations and risking the exacerbation  
360 of health disparities<sup>43,44</sup>. Consistent with prior research, our study observed that a lung cancer  
361 PRS constructed using EUR data significantly underperformed when applied to EAS individuals<sup>15</sup>.  
362 However, by utilizing novel methods to integrate GWAS data from both EUR and EAS  
363 populations, we demonstrated that a multi-ancestry PRS enhances the precision of risk  
364 stratification for LUAD among EAS never-smokers.

365

366 While large-scale efforts to integrate PRS into clinical practice have focused on conditions with  
367 well-established PRS prediction performance<sup>16-18</sup>, such as Type 2 Diabetes, breast cancer, and  
368 cardiovascular diseases, lung cancer has been notably absent from these efforts, primarily due  
369 to the overwhelming focus on smoking as a risk factor. This has limited the exploration and  
370 application of PRS in lung cancer risk stratification, particularly among never-smokers. Our study  
371 addresses this gap by presenting a PRS that signifies a 3.92-fold increase in lung cancer risk for  
372 individuals in the top 20% risk quantile, a marked improvement over the 2.09-fold increase  
373 observed in the prior study by Wei et al.<sup>45</sup> for a cohort of Chinese never-smoking women.

374

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

383

384 Current PRS studies predominantly focus on evaluating relative and absolute risks<sup>8,12,15,45,46</sup>, yet  
385 our research introduces a crucial forecast of sample sizes required for differing levels of PRS  
386 accuracy. Our projection analyses indicated that achieving an AUC of 0.70 would approach the  
387 maximized prediction potential of the PRS, given the estimated genetic variance explained by  
388 GWAS chip variants. To reach this level of accuracy, a future study would need to include 65,000  
389 EAS LUAD cases with a 1:4 case-control ratio, or 55,000 cases with a 1:10 case-control ratio.  
390 Notably, accessing large biobanks with publicly available controls could reduce the number of  
391 required cases<sup>47</sup>. These projections offer a strategic framework for planning and designing  
392 future genetic studies on lung cancer, establishing clear benchmarks for PRS model  
393 performance, which is essential for the scientific community in advancing genetic risk  
394 prediction.

395

396 Our study has several limitations. First, it focuses on genetic susceptibility without considering  
397 additional questionnaire-based risk factors, aiming primarily to identify the most effective PRS  
398 model within genetic contexts. We plan to conduct integrative analyses that merge PRS with  
399 other risk indicators to refine LUAD risk predictions for never-smokers. Second, our projections  
400 are tailored to single-ancestry PRS models, not fully addressing the intricacies of multi-ancestry  
401 approaches, including diverse sample sizes and genetic correlations across populations. Third,  
402 our validation cohort, derived from the FLCCA and covering various EAS regions, relied on  
403 Taiwanese incidence data for absolute risk estimations. Consequently, our absolute risk findings  
404 may not extend universally to all EAS never-smokers, though our PRS performance assessment  
405 and sample size projections remain applicable to our population of interest. Lastly, the clinical  
406 application of our PRS findings, particularly in screening and risk counseling, is still in  
407 preliminary stages. Despite advancements in using PRS for smoking cessation trials<sup>48</sup>, its  
408 implementation for never-smokers in clinical settings has been limited, presenting a significant  
409 area for future research.

410

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

419

420

421

422

## 423 **Acknowledgements**

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

426 Female Lung Cancer Consortium in Asia (NCI): This study was supported by a Grant-in-Aid for  
427 Scientific Research on Priority Areas from the Ministry of Education, Science, Sports, Culture  
428 and Technology of Japan, a Grant-in- Aid for the Third Term Comprehensive 10-Year Strategy for  
429 Cancer Control from the Ministry Health, Labor and Welfare of Japan, by Health and Labor  
430 Sciences Research Grants for Research on Applying Health Technology from the Ministry of  
431 Health, Labor and Welfare of Japan, by the National Cancer Center Research and Development  
432 Fund, the National Research Foundation of Korea (NRF) grant funded by the Korea government  
433 (MEST) (grant No. 2011-0016106), a grant of the National Project for Personalized Genomic  
434 Medicine, Ministry for Health & Welfare, Republic of Korea (A111218-11-GM04), the Program  
435 for Changjiang Scholars and Innovative Research Team in University in China (IRT\_14R40 to  
436 K.C.), the National Science & Technology Pillar Program (2011BAI09B00), MOE 111 Project  
437 (B13016), the National Natural Science Foundation of China (No. 30772531, and 81272618),  
438 Guangdong Provincial Key Laboratory of Lung Cancer Translational Medicine (No.  
439 2012A061400006), Special Fund for Research in the Public Interest from the National Health  
440 and Family Planning Commission of PRC (No. 201402031), and the Ministry of Science and  
441 Technology, Taiwan (MOST 103-2325-B-400-023 & 104-2325-B-400-012). The Japan Lung Cancer  
442 Study (JLCS) was supported in part by the Practical Research for Innovative Cancer Control from  
443 Japan Agency for Medical Research and Development (15ck0106096h0002) and the  
444 Management Expenses Grants from the Government to the National Cancer Center (26-A-1) for  
445 Biobank. BioBank Japan was supported by the Ministry of Education, Culture, Sports, Sciences  
446 and Technology of the Japanese government. The Japan Public Health Center-based prospective  
447 Study (the JPHC Study) was supported by the National Cancer Center Research and  
448 Development Fund (23-A- 31[toku], 26-A-2, 29-A-4, 2020-J-4, and 2023-J-4) (since 2011) and a  
449 Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan  
450 (from 1989 to 2010). The Taiwan GELAC Study (Genetic Epidemiological Study for Lung  
451 AdenoCarcinoma) was sup- ported by grants from the National Research Program on Genomic



452 Medicine in Taiwan (DOH99-TD-G-111-028), the National Research Program for  
453 Biopharmaceuticals in Taiwan (MOHW 103-TDUPB-211-144003, MOST 103-2325-B-400-023)  
454 and the Bioinformatics Core Facility for Translational Medicine and Biotechnology Development  
455 (MOST 104-2319-B-400-002). This work was also supported by the Jinan Science Research  
456 Project Foundation (201102051), the National Key Scientific and Technological Project  
457 (2011ZX09307-001-04), the National Natural Science Foundation of China (No.81272293), the  
458 State Key Program of National Natural Science of China (81230067), the National Research  
459 Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-  
460 2014R1A2A2A05003665), Sookmyung Women's University Research Grants, Korea (1-1603-  
461 2048), Agency for Science, Technology and Research (A\*STAR), Singapore and the US National  
462 Institute of Health Grant (1U19CA148127-01). The overall GWAS project was supported by the  
463 intramural program of the US National Institutes of Health/National Cancer Institute. The  
464 following is a list of grants by study center: SKLCS (Y.T.K.)—National Research Foundation of  
465 Korea (NRF) grant funded by the Korea government (MEST) (2011-0016106). (J.C.) – This work  
466 was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health  
467 & Welfare, Republic of Korea (grant no. 0720550-2). (J.S.S) – grant number is A010250. WLCS  
468 (T.W.)—National Key Basic Research and Development Program (2011CB503800). SLCS (B.Z.)—  
469 National Nature Science Foundation of China (81102194). Liaoning Provincial Department of  
470 Education (LS2010168). China Medical Board (00726). GDS (Y.L.W.)—Foundation of Guangdong  
471 Science and Technology Department (2006B60101010, 2007A032000002, 2011A030400010).  
472 Guangzhou Science and Information Technology Bureau (2011Y2-00014). Chinese Lung Cancer  
473 Research Foundation, National Natural Science Foundation of China (81101549). Natural  
474 Science Foundation of Guangdong Province (S2011010000792). TLCS (K.C., B.Q)—Program for  
475 Changjiang Scholars and Innovative Research Team in University (PCSIRT), China (IRT1076).  
476 Tianjin Cancer Institute and Hospital. National Foundation for Cancer Research (US). FLCS  
477 (J.C.W., D.R., L.J.)—Ministry of Health (201002007). Ministry of Science and Technology  
478 (2011BAI09B00). National S&T Major Special Project (2011ZX09102-010-01). China National  
479 High-Tech Research and Development Program (2012AA02A517, 2012AA02A518). National  
480 Science Foundation of China (30890034). National Basic Research Program (2012CB944600).

481 Scientific and Technological Support Plans from Jiangsu Province (BE2010715). NLCS (H.S.)—  
482 China National High-Tech Research and Development Program Grant (2009AA022705). Priority  
483 Academic Program Development of Jiangsu Higher Education Institution. National Key Basic  
484 Research Program Grant (2011CB503805). GEL-S (A.S.)—National Medical Research Council  
485 Singapore grant (NMRC/0897/2004, NMRC/1075/2006). (J.Liu)—Agency for Science, Technology  
486 and Research (A\*STAR) of Singapore. GELAC (C.A.H.)—National Research Program on Genomic  
487 Medicine in Taiwan (DOH98-TDG-111-015). National Research Program for Biopharmaceuticals  
488 in Taiwan (DOH 100- TD-PB-111-TM013). National Science Council, Taiwan (NSC 100- 2319-B-  
489 400-001). YLCS (Q.L.)—Supported by the intramural pro- gram of U.S. National Institutes of  
490 Health, National Cancer Institute. SWHS (W.Z., W-H.C., N.R.)—The work was supported by a  
491 grant from the National Institutes of Health (R37 CA70867, UM1 CA182910) and the National  
492 Cancer Institute intramural research program, including NCI Intramural Research Program  
493 contract (N02 CP1101066). JLCS (K.M., T.K.)—Grants-in-Aid from the Ministry of Health, Labor,  
494 and Welfare for Research on Applying Health Technology and for the 3rd-term Comprehensive  
495 10-year Strategy for Cancer Control; by the National Cancer Center Research and Development  
496 Fund; by Grant-in-Aid for Scientific Research on Priority Areas and on Innovative Area from the  
497 Ministry of Education, Science, Sports, Culture and — Technology of Japan. (W.P.)—NCI R01-  
498 CA121210. HKS (J.W.)— General Research Fund of Research Grant Council, Hong Kong  
499 (781511M). The Environment and Genetics in Lung Cancer Etiology (EAGLE), Prostate, Lung,  
500 Colon, Ovary Screening Trial (PLCO), and Alpha-Tocopherol, Beta-Carotene Cancer Prevention  
501 (ATBC) studies were supported by the Intramural Research Program of the National Institutes of  
502 Health, National Cancer Institute (NCI), Division of Cancer Epidemiology and Genetics. ATBC was  
503 also supported by U.S. Public Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-  
504 RC-37004) from the NCI. PLCO was also supported by individual contracts from the NCI to the  
505 University of Colorado Denver (NO1-CN-25514), Georgetown University (NO1-CN-25522), the  
506 Pacific Health Research Institute (NO1-CN-25515), the Henry Ford Health System (NO1-CN-  
507 25512), the University of Minnesota, (NO1-CN- 25513), Washington University (NO1-CN-25516),  
508 the University of Pittsburgh (NO1-CN-25511), the University of Utah (NO1-CN- 25524), the  
509 Marshfield Clinic Research Foundation (NO1-CN- 25518), the University of Alabama at

510 Birmingham (NO1-CN- 75022), Westat, Inc. (NO1-CN-25476), and the University of California,  
511 Los Angeles (NO1-CN-25404). The Carotene and Retinol Efficacy Trial (CARET) is funded by the  
512 National Cancer Institute, National Institutes of Health through grants U01-CA063673, UM1-  
513 CA167462, and U01-CA167462. The Cancer Prevention Study-II (CPS-II) Nutrition Cohort was  
514 supported by the American Cancer Society. The NIH Genes, Environment and Health Initiative  
515 (GEI) partly funded DNA extraction and statis- tical analyses (HG-06-033-NCI-01 and  
516 RO1HL091172-01), genotyping at the Johns Hopkins University Center for Inherited Disease  
517 Research. This research was supported by the National Research Foundation of Korea (NRF)  
518 grant funded by the Korea government (MSIT) (No. 2020R1A2C4002236)

519

520

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

523

524 Female Lung Cancer Consortium in Asia (Taiwan): The Ministry of Health and Welfare grants  
525 DOH97-TD-G-111-028 (ISC), DOH98-TD-G-111-017 (ISC), DOH99-TD-G-111-014 (ISC); DOH97-  
526 TD-G-111-026 (CAH), DOH98-TD-G-111-015 (CAH), DOH99-TD-G-111-028 (CAH); National  
527 Health Research Institutes grants NHRI-PH-110-GP-01, NHRI-PH-110-GP-03; and the Ministry of  
528 Science and Technology grants MOST108-2314-B-400-038(CAH), MOST109-2740-B-400-  
529 002(CAH), MOST 111-2740-B-400-002 (CAH), MOST 111-2314-B-400-020 (CAH).

530

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

533

534 OncoArray study in Europeans: The OncoArray data and analysis from INTEGRAL-ILCCO were  
535 supported by NIH U19 CA203654, and U19 CA148127. The data harmonization for ILCCO was  
536 supported by Canadian Institute for Health Research (CIHR) Canada Research Chair to R.J.H, and  
537 CIHR FDN 167273).

538

539 European never-smoking lung cancer study: CIA is a Research Scholar of the Cancer Prevention  
540 Institute of Texas (CPRIT) and supported by CPRIT grant RR170048.

541

542 Taiwan eQTL study: This study was supported by the Ministry of Health and Welfare grants  
543 DOH97-TD-G-111-028 (ISC), DOH98-TD-G-111-017 (ISC), DOH99-TD-G-111-014 (ISC); DOH97-  
544 TD-G-111-026 (CAH), DOH98-TD-G-111-015 (CAH), DOH99-TD-G-111-028 (CAH); National  
545 Health Research Institutes grants NHRI-PH-110-GP-01, NHRI-PH-110-GP-03; and the Ministry of  
546 Science and Technology grants MOST108-2314-B-400-038(CAH), MOST109-2740-B-400-  
547 002(CAH).

548 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  
549 supported by NIH grant 1R01HG010480. P.Y. is supported by Mayo Clinic Foundation Research  
550 Funds, NIH-CA77118 and CA80127. G.L. is supported is supported by the Alan Brown Chair and  
551 Lusi Wong Fund of the Princess Margaret Cancer Foundation. D.C.C is supported by  
552 U01CA209414. O.Y.G. is supported by NIH R01 CA231141.

553

554

#### 555 **Author contributions**

556 B.B., N.R., Q.L. and H.Z. organized and designed the study. B.B., K.W., H.Z. carried out the  
557 statistical analyses. B.B., K.W., Q.L. N.R. and H.Z. wrote the first draft. Q.L., M.T.L., B.A.B., W.H.,  
558 N.E.C., BT.J., M.Song, H.P., D.A., C.C.C., L.B., M.Y., A.H., B.H., J.Liu, B.Zhu, S.I.B., C.K., K.Wyatt,  
559 S.A.L., A.Chao, J.F.F.J., S.J.C., N.R., Z.Wang, C..L., J.C., C.W., W.T., D.Lin, S.J.A., X.C.Z., J.S., YL.W.,  
560 M.P.W., L.P.C., J.C.M.H., V.H.F.L., Z.H., K.M., J.Y..P., Jia.Liu, HS.J., J.E..C., Y.Y.C., H.N.K., MH.S., SS.K.,  
561 YC.K., IJ.O., S.W.S., HI.Y., Y.T.K., YC.H., J.H.K., Y.H.K., J.S..S., Y.J..J., K.H.P., C.H.K., J.S.K., I.K.P., B..S.,  
562 Jie.L, Z.W., S.C., J.Y., J.C.W., Y.Y., YB.X., YT.G., D.L., J.Y.Y.W., H.C., L..J., J.Z., G.J., K..F., Z.Y., B.Z., W.W.,  
563 P.G., Q.H., X.L., Y.R., A.S., Y.L., Y.C., WY.L., W.Z., XO.S., Q.C., G.Y., B.Q., T.W., H.G., L..L., P.X., F.W.,  
564 G.W., J.X., J.L., R.CH.V., B.B., H.D.III.H., J.Wang, A.D.L..S., J.KC.C., V.L.S., K.C., H.Z., H.D., C.A.H.,  
565 TY.C., LH.C..., IS.C., CY.C., S.S.J ., CH.C., GC.C., CF.H., YH.T., WC.W., KY.C., MS.H., WC.S., YM.C.,  
566 CL.W., KC.C., CJ.Y., HH.H., FY.T., HC.L., CJ.C., PC.Y., K.Shiraishi, T.K., H.K., S.M., H.H., K.Goto, Y.Ohe,  
567 S.W., Y.Yatabe, M.T., R.Hamamoto, A.Takahashi, Y.Momozawa, M.Kubo, Y.K., Y.D., Y.Miyagi, H.N.,

568 T.Y., N.S., M.I., M.H., Y.N., K.Takeuchi, K.W., K.Matsuda, Y.Murakami, K.S., K.T., Y.O., M.S.,  
569 H.Suzuki, A.G., Y.M., T.H., M.K., K.O., H.S., J.D., H.M., M.Z., R.J.H., S.L., A.T., C.C., S.E.B.,  
570 M.Johansson, A.R., H.Bö., HE.W., D.C., G.R., S.A., P.B., J.MK., J.K.F., S.S.S., L.L.M., O.M., H.Bö.,  
571 G.L., A.A., L.A.K., S.ZN., K.G., M.J., A.C., JM.Y., P.L., M.B.S., M.C.A., C.I.A., A.G.S., R.H., M.R.S.,  
572 O.Y.G., I.P.G., X.W., P.Y. conducted epidemiology studies and contributed samples to GWAS  
573 and/or conducted initial genotyping. All authors reviewed and approved the final review of the  
574 manuscript.

575

#### 576 **Competing interests**

577 The authors declare no competing interests.

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593 **Table 1. Prediction performance of different methods for generating polygenic risk scores for**  
 594 **lung cancer in never-smoking East Asian populations.**

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

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

597 <sup>1</sup>ORs are provided per one unit standard deviation of PRS.

598 <sup>2</sup>Adjusted for age (continuous) and first 10 principal components.

599 <sup>3</sup>Sample size for single-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian  
 600 (EAS) ancestry for training, 2,219 cases and 2,272 controls of EAS ancestry for tuning and 2,219 cases  
 601 and 2,272 controls of EAS ancestry for validation.

602 <sup>4</sup>SNPs that reached genome-wide significance in Shi et al., 2023.

603 <sup>5</sup>SNPs obtained from genome-wide association study conducted in individuals with European ancestry  
 604 (Hung et al., 2021, Cancer Res.).

605 <sup>6</sup>Sample size for multi-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian  
 606 (EAS) ancestry, as well as 2,058 cases and 5,575 controls of European ancestry for training, 2,219 cases  
 607 and 2,272 controls of EAS ancestry for tuning and 2,219 cases and 2,272 controls of EAS ancestry for  
 608 validation.

609

610 **Table 2. Genetic variance in East Asian lung cancer among never smokers explained by**  
 611 **LDpred2 PRS**

Sample size <sup>1</sup>	Genetic variance of PRS <sup>2</sup>	Proportion of all-GWAS variants genetic variance explained by PRS <sup>3</sup>	Proportion of Familial risk explained by PRS <sup>4</sup>
Current	0.201	26.6%	16.5%
35,000 cases	0.437	57.9%	35.8%
55,000 cases	0.540	71.5%	44.3%

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

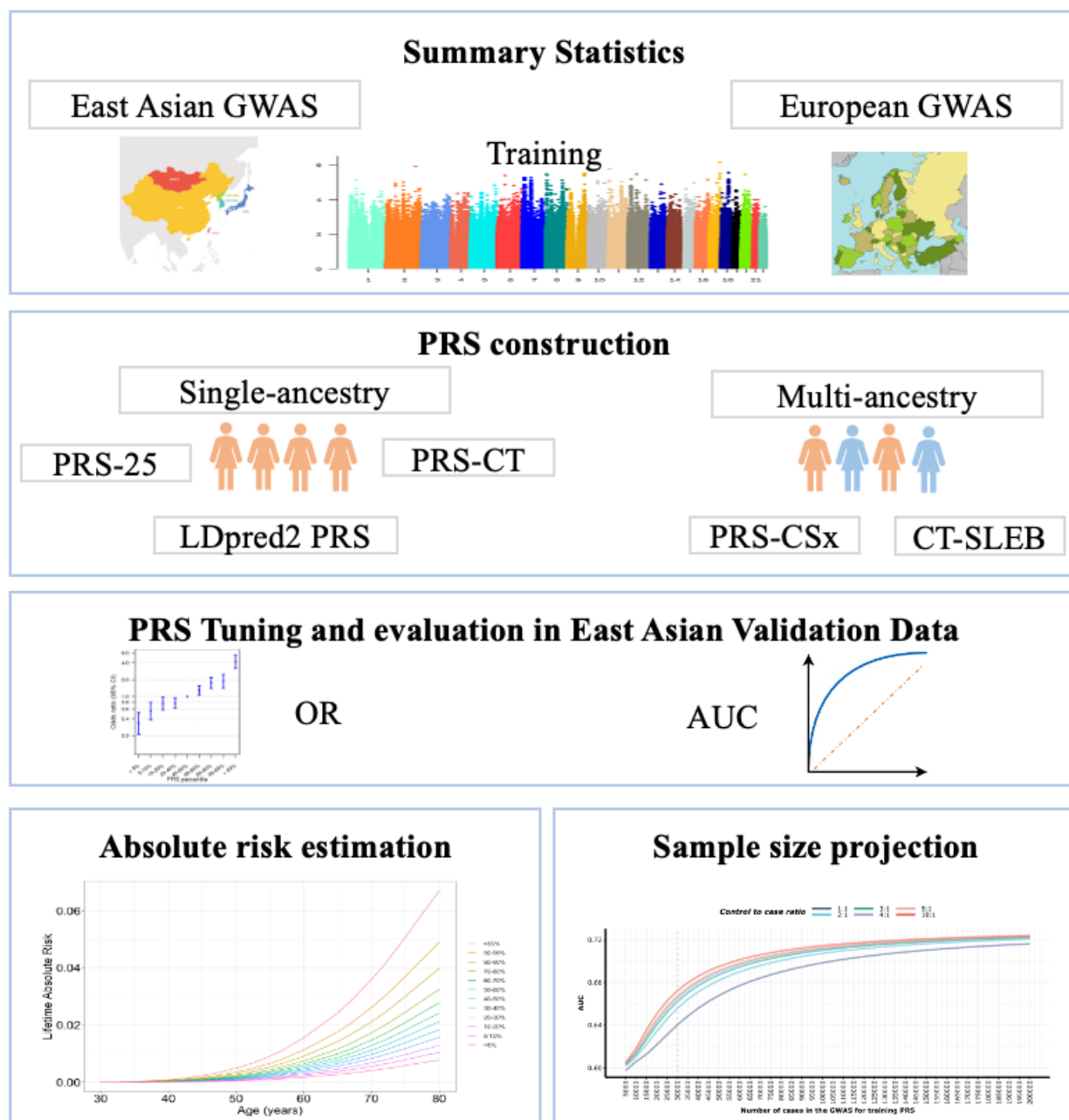
617 <sup>2</sup> *Genetic Variance of PRS*: This corresponds to the heritability on the frailty scale, assuming a polygenic  
 618 log-additive model underpins this relationship. It quantifies the proportion of the phenotype variation  
 619 that can be attributed to genetic factors in the context of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

620 <sup>3</sup> *Proportion of Genetic Variance from All-GWAS Variants Explained by PRS*: This represents the variance  
 621 of all genome-wide imputable variants as established through LD-score regression (refer to Nat. Genet.  
 622 47, 291-5 (2015) and Nat. Genet. 47, 1236-41 (2015)). On the frailty scale, the genetic variance of all  
 623 GWAS variants is calculated as  $\sigma_{GWAS}^2 = Var(\sum_{m=1}^M \beta_m G_m)$ , where  $G_m$  is the standardized genotype for  
 624 the mth SNP,  $\beta_m$  is the true log odds ratio for the mth SNP and M is the total number of causal SNPs  
 625 among the GWAS variants. For East Asian never smokers, the estimated genetic variance of all GWAS  
 626 variants is 0.755.

627 <sup>4</sup> *Proportion of Familial Risk Explained by PRS*: This calculates the familial risk in terms of genetic variance  
 628 using the formula  $\lambda_s^2 = \exp(\sigma^2)$ , where  $\lambda_s$  is the familial risk when a first-order sibling has the disease,  
 629 and  $\sigma^2$  is the genetic variance on frailty-scale. Further details of this calculation can be found in Nat.  
 630 Genet. 31, 33-36 (2002). For lung cancer in East Asian never-smokers, the familial risk is a 1.84-fold  
 631 increase, and the genetic variance of all GWAS variants, as estimated through LD-score regression,  
 632 explains 61.9% of this increased familial relative risk.

633

634



635

636 **Figure 1. Overview of data structure, polygenic risk score (PRS) development, validation and**

637 **application.** Summary statistics from East Asian (EAS) genome-wide association studies (GWAS)

638 were used to develop single-ancestry PRS using methods such as a simple PRS constructed using

639 25 SNPs that have previously reached genome-wide significance (i.e.,  $P < 5 \times 10^{-8}$ ) (PRS-25), a PRS

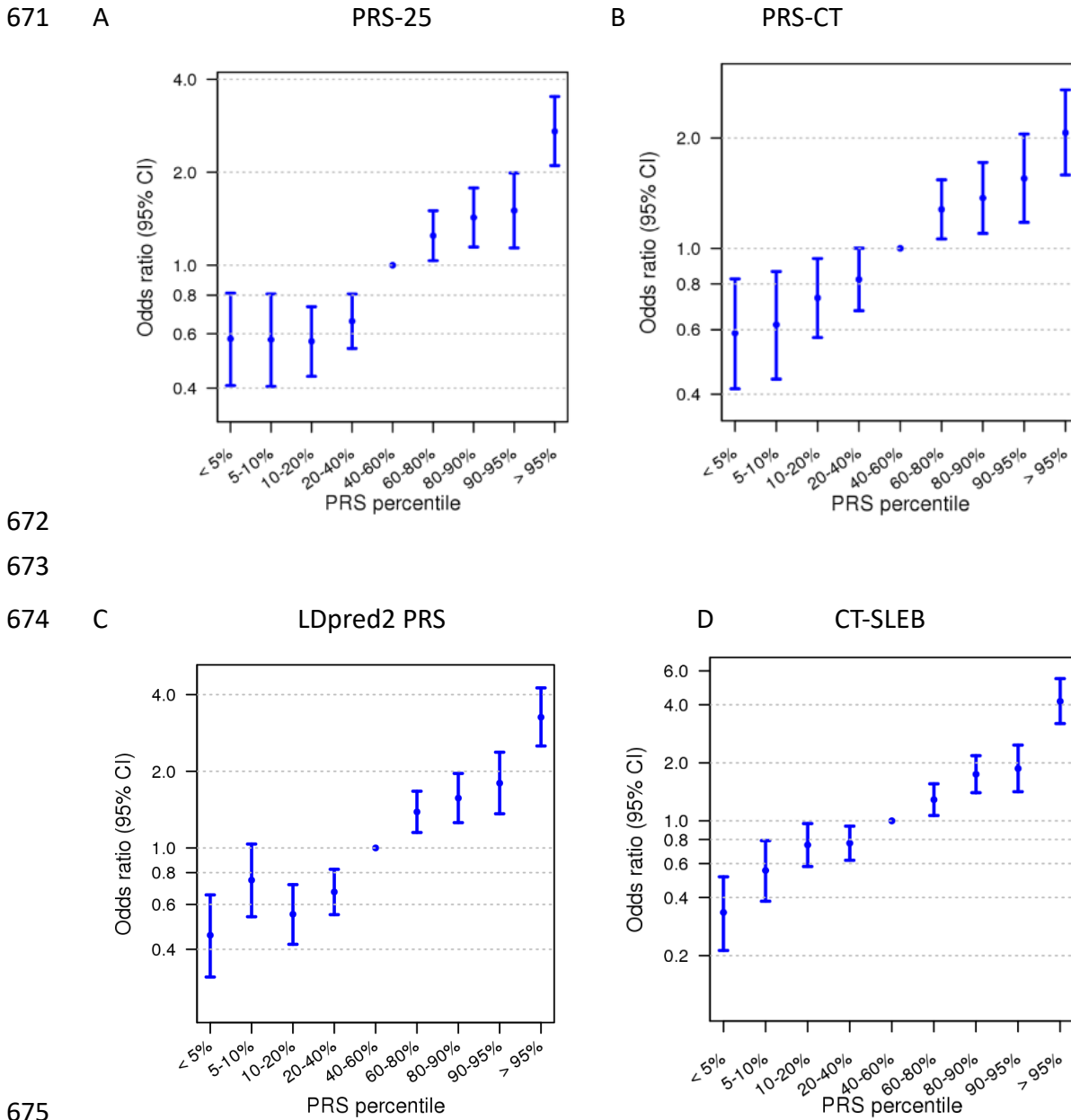
640 using the clumping and thresholding (CT) method (PRS-CT) incorporating 8 SNPs, and a PRS

641 using a genome-wide Bayesian-based approach, LDpred2 (LDpred2 PRS) incorporating close to a



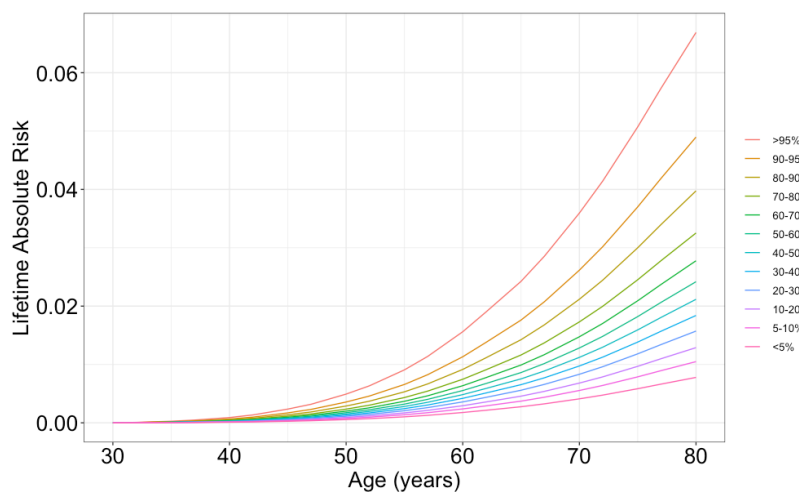
642 million SNPs. For the multi-ancestry PRS development, we also used summary statistics from  
643 European (EUR) GWAS, applying the PRS-CSx method that leveraged genome-wide association  
644 summary statistics for close to a million SNPs with a Bayesian continuous shrinkage prior to  
645 model SNP effect sizes across populations, as well as CT-SLEB method, which enhances the  
646 standard CT methods with a two-dimensional approach to select SNPs for EAS PRS construction  
647 by incorporating over 2 million SNPs. Tuning and validation of each PRS was conducted in an  
648 independent EAS individual-level data. Relative risk per PRS quantile was calculated as an odds  
649 ratio (OR) with the middle quantile (40<sup>th</sup> to 60<sup>th</sup> percentile) set as the reference, and the area  
650 under the receiver operating curve (AUC) was estimated for each PRS. CT-SLEB PRS was used to  
651 estimate 10-year and lifetime cumulative absolute risk, and PRS-CT and PRS-LDpred2 were used  
652 for sample size projection.

653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670

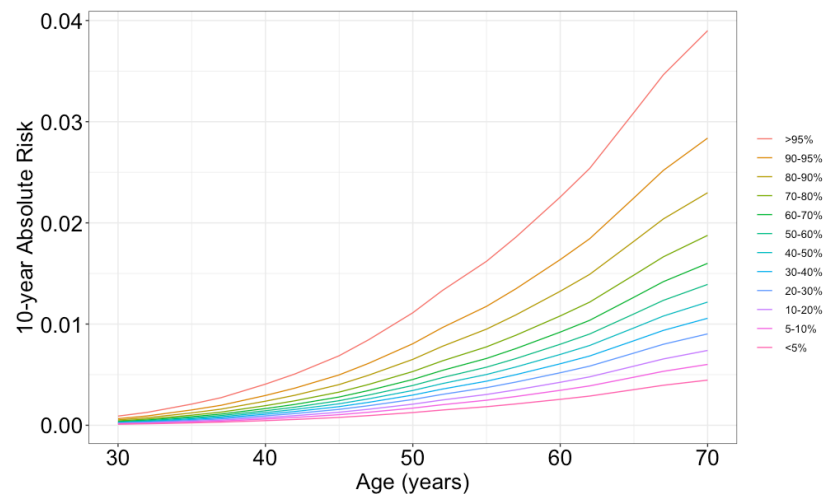


678 **Figure 2. Relative risk estimated for quantiles of each polygenic risk score (PRS) and lung**  
679 **adenocarcinoma in the validation dataset of women with East Asian ancestry, treating the**  
680 **40<sup>th</sup> to 60<sup>th</sup> percentile as the references.** Odds ratios of PRS per standard deviation (SD) and  
681 95% confidence intervals are shown for the single-ancestry 25 SNP polygenic risk score, PRS-25  
682 (A), Clumping and thresholding method, PRS-CT (B), Bayesian-based genome-wide approach,  
683 LDpred2 PRS (C), multi-ancestry approach, CT-SLEB (D).

A

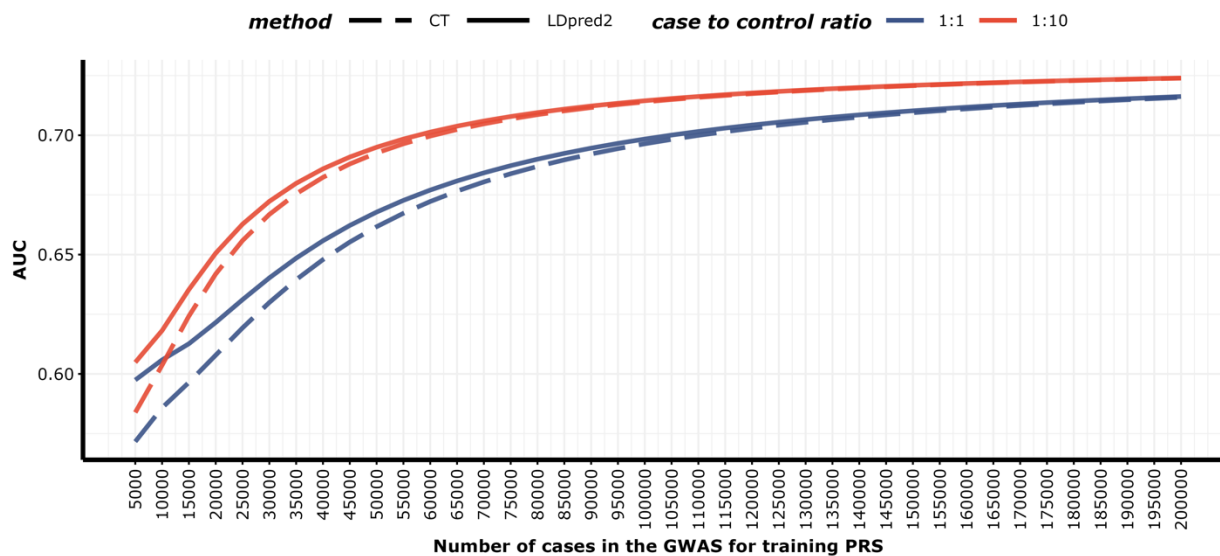


B

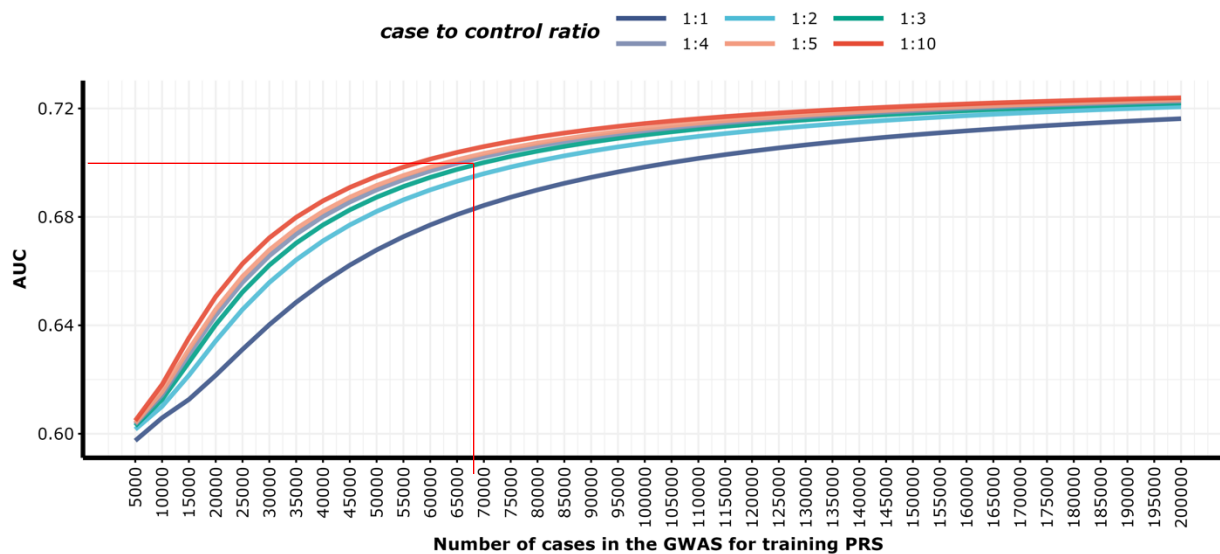


**Figure 3. Lifetime cumulative and 10-year absolute risk of developing lung cancer.** Lifetime Cumulative (A) 10-year (B) absolute risk of developing lung adenocarcinoma in never-smoking women in East Asia by percentiles of the CT-SLEB polygenic risk score (PRS). Absolute risks were calculated based on Taiwan's age specific incidence and mortality data and using the PRS relative risks estimated as described in the material and methods section.

A



B



**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

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
2. Sun, S., Schiller, J. H. & Gazdar, A. F. Lung cancer in never smokers--a different disease. *Nat. Rev. Cancer* **7**, 778–790 (2007).
3. Cheng, T. Y. D. *et al.* The International Epidemiology of Lung Cancer: Latest Trends, Disparities, and Tumor Characteristics. *J. Thorac. Oncol.* **11**, 1653–1671 (2016).
4. Byun, J. *et al.* ‘Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237 controls identifies new susceptibility loci contributing to lung cancer’. *Nat Genet* **54**, 1167 (2022).
5. Landi, M. T. *et al.* Tracing lung cancer risk factors through mutational signatures in never-smokers. *Am. J. Epidemiol.* **190**, 962–976 (2021).
6. Corrales, L. *et al.* Lung cancer in never smokers: The role of different risk factors other than tobacco smoking. *Crit. Rev. Onco. Hematol.* **148**, (2020).
7. Sisti, J. & Boffetta, P. What proportion of lung cancer in never-smokers can be attributed to known risk factors? *Int. J. Cancer* **131**, 265–275 (2012).
8. Shi, J. *et al.* Genome-wide association study of lung adenocarcinoma in East Asia and comparison with a European population. *Nature Commun.* **14**, (2023).
9. Dai, J. *et al.* Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lanet. Respir. Med.* **7**, 881–891 (2019).
10. Seow, W. J. *et al.* Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from Western populations. *Hum. Mol. Genet.* **26**, 454–465 (2017).
11. Lan, Q. *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat. Genet.* **44**, 1330–1335 (2012).
12. Li, Y. *et al.* Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lanet Oncol.* **11**, 321–330 (2010).
13. Hung, R. J. *et al.* Lung cancer risk in never-smokers of European descent is associated with genetic variation in the 5p15.33 TERT-CLPTM11L region. *J. Thorac Oncol.* **14**, 1360–1369 (2019).
14. Lebrecht, M. B. *et al.* Validation of lung cancer polygenic risk scores in a high-risk case-control cohort. *Genet. Med.* **25**, (2023).
15. Hung, R. J. *et al.* Assessing Lung Cancer Absolute Risk Trajectory Based on a Polygenic Risk Model. *Cancer Res.* **81**, 1607–1615 (2021).
16. Linder, J. E. *et al.* Returning integrated genomic risk and clinical recommendations: The eMERGE study. *Genet. Med.* **25**, (2023).
17. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting workflow. *Nat. Med.* **28**, 1006–1013 (2022).
18. Shieh, Y. *et al.* Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J. Natl. Cancer Inst.* **109**, (2017).

19. Kachuri, L. *et al.* Principles and methods for transferring polygenic risk scores across global populations. *Nat. Rev. Genet.* **25**, 8–25 (2024).
20. Jin, J. *et al.* MUSSEL: Enhanced Bayesian polygenic risk prediction leveraging information across multiple ancestry groups. *Cell Genom.* **4**, (2024).
21. Zhang, J. *et al.* An ensemble penalized regression method for multi-ancestry polygenic risk prediction. *Nat. Commun.* **15**, 1–14 (2024).
22. Hou, K. *et al.* Admix-kit: an integrated toolkit and pipeline for genetic analyses of admixed populations. *Bioinformatics* **40**, (2024).
23. Hoggart, C. J. *et al.* BridgePRS leverages shared genetic effects across ancestries to increase polygenic risk score portability. *Nat. Genet.* **56**, 180–186 (2024).
24. Zhang, H. *et al.* A new method for multi-ancestry polygenic prediction improves performance across diverse populations. *Nat. Genet.* **55**, 1757–1768 (2023).
25. Zhou, G., Chen, T. & Zhao, H. SDPRX: A statistical method for cross-population prediction of complex traits. *Am. J. Hum. Genet.* **110**, 13–22 (2023).
26. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
27. Shi, J. *et al.* Genome-wide association study of lung adenocarcinoma in East Asia and comparison with a European population. *Nature Communications* **14**:1 **14**, 1–17 (2023).
28. Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.* **43**, 792–796 (2011).
29. Wang, L. *et al.* Genetically determined height was associated with lung cancer risk in East Asian population. *Cancer Med.* **7**, 3445–3452 (2018).
30. Dong, J. *et al.* Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat. Genet.* **44**, 895–898 (2012).
31. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
32. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
33. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).
34. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
35. Hung, R. J. *et al.* Assessing Lung Cancer Absolute Risk Trajectory based on a Polygenic Risk Model. *Cancer Res* **81**, 1607 (2021).
36. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
37. Foucher, Y. *et al.* RISCA: Causal Inference and Prediction in Cohort-Based Analyses. *R package version 1.01* Preprint at <https://cran.r-project.org/package=RISCA> (2020).
38. Choudhury, P. P. *et al.* iCARE: An R package to build, validate and apply absolute risk models. *PLoS One* **15**, (2020).
39. Chien, L. H. *et al.* Predicting Lung Cancer Occurrence in Never-Smoking Females in Asia: TNSF-SQ, a Prediction Model. *Cancer Epidemiol. Biomarkers Prev.* **29**, 452–459 (2020).

40. Zhang, Y., Qi, G., Park, J. H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).
41. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
42. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, (2019).
43. Kullo, I. J. *et al.* Polygenic scores in biomedical research. *Nat. Rev. Genet.* **23**, 524–532 (2022).
44. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
45. Wei, X. *et al.* Development and evaluation of a polygenic risk score for lung cancer in never-smoking women: A large-scale prospective Chinese cohort study. *Int. J. Cancer* **154**, 807–815 (2024).
46. Blechter, B. *et al.* Polygenic risk score, environmental tobacco smoke, and risk of lung adenocarcinoma in never-smoking women in Taiwan. *JAMA Netw. Open* **6**, E2339254 (2023).
47. Katki, H. A. *et al.* Increase in power by obtaining 10 or more controls per case when type-1 error is small in large-scale association studies. *BMC Med. Res. Methodol.* **23**, (2023).
48. Chen, T. *et al.* Genomic Insights for Personalized Care: Motivating At-Risk Individuals Toward Evidence-Based Health Practices. *medRxiv* 2024.03.19.24304556 (2024) doi:10.1101/2024.03.19.24304556.

## 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](mailto:kshirais@ncc.go.jp) or Takashi Kohno at [tkkohno@ncc.go.jp](mailto: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](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000876.v2.p1)), phs001273.v3.p2 (Oncoarray Consortium, [32](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-</a></p></div><div data-bbox=)



[bin/study.cgi?study\\_id=phs001273.v3.p2](#)). 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

Methods implemented and their corresponding repositories include: SCT and LDpred2 at <https://github.com/privefl/bigsnpr>, PRS-CSx at <https://github.com/getian107/PRScsx>. CT-SLEB at <https://github.com/andrewhaoyu/CTSLEB>, and GENESIS at <https://github.com/yandorazhang/GENESIS>.

PLINK: <https://www.cog-genomics.org/plink/1.9>.

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](https://github.com/KevinWFred/PRS_EASLC)

## Never-Smoker Lung Cancer Working Group

Baosen Zhou<sup>1</sup>, Jiucun Wang<sup>2,3</sup>, Wei Jie Seow<sup>4,5,6</sup>, Lei Song<sup>4</sup>, Wei Hu<sup>4</sup>, Li-Hsin Chien<sup>7,8</sup>, Qiuyin Cai<sup>9</sup>, Hee Nam Kim<sup>10</sup>, Yi-Long Wu<sup>11</sup>, Maria Pik Wong<sup>12</sup>, Brian Douglas Richardson<sup>4,13</sup>, Shilan Li<sup>4,14</sup>, Tongwu Zhang<sup>4</sup>, Charles Breeze<sup>4</sup>, Zhaoming Wang<sup>15</sup>, Bryan A Bassig<sup>16</sup>, Jin Hee Kim<sup>17</sup>, Demetrius Albanes<sup>4</sup>, Jason YY Wong<sup>18</sup>, Min-Ho Shin<sup>19</sup>, Lap Ping Chung<sup>12</sup>, Yang Yang<sup>20</sup>, She-Juan An<sup>11</sup>, Hong Zheng<sup>21</sup>, Yasushi Yatabe<sup>22</sup>, Xu-Chao Zhang<sup>11</sup>, Young-Chul Kim<sup>23,24</sup>, Neil E Caporaso<sup>4</sup>, Jiang Chang<sup>25</sup>, James Chung Man Ho<sup>26</sup>, Michiaki Kubo<sup>27</sup>, Yataro Daigo<sup>28,29</sup>, Minsun Song<sup>30</sup>, Yukihide Momozawa<sup>27</sup>, Yoichiro Kamatani<sup>31</sup>, Masashi Kobayashi<sup>32</sup>, Kenichi Okubo<sup>32</sup>, Takayuki Honda<sup>33</sup>, H Dean Hosgood<sup>34</sup>, Hideo Kunitoh<sup>35</sup>, Shun-ichi Watanabe<sup>36</sup>, Yohei Miyagi<sup>37</sup>, Haruhiko Nakayama<sup>38</sup>, Shingo Matsumoto<sup>39</sup>, Hidehito Horinouchi<sup>36</sup>, Masahiro Tsuboi<sup>40</sup>, Ryuji Hamamoto<sup>41</sup>, Koichi Goto<sup>39</sup>, Yuichiro Ohe<sup>36</sup>, Atsushi Takahashi<sup>31</sup>, Akiteru Goto<sup>42</sup>, Yoshihiro Minamiya<sup>43</sup>, Megumi Hara<sup>44</sup>, Yuichiro Nishida<sup>44</sup>, Kenji Takeuchi<sup>45</sup>, Kenji Wakai<sup>45</sup>, Koichi Matsuda<sup>46</sup>, Yoshinori Murakami<sup>47</sup>, Kimihiro Shimizu<sup>48</sup>, Hiroyuki Suzuki<sup>49</sup>, Motonobu Saito<sup>50</sup>, Yoichi Ohtaki<sup>51</sup>, Kazumi

Tanaka<sup>51</sup>, Tangchun Wu<sup>52</sup>, Fusheng Wei<sup>53</sup>, Hongji Dai<sup>21</sup>, Mitchell J Machiela<sup>4</sup>, Jian Su<sup>11</sup>, Yeul Hong Kim<sup>54</sup>, In-Jae Oh<sup>23,24</sup>, Victor Ho Fun Lee<sup>55</sup>, Gee-Chen Chang<sup>56,57,58,59</sup>, Ying-Huang Tsai<sup>60,61</sup>, Kuan-Yu Che<sup>62</sup>, Ming-Shyan Huang<sup>63</sup>, Wu-Chou Su<sup>64</sup>, Yuh-Min Chen<sup>65</sup>, Adeline Seow<sup>5</sup>, Jae Yong Park<sup>66</sup>, Sun-Seog Kweon<sup>19,67</sup>, Kun-Chieh Chen<sup>57</sup>, Yu-Tang Gao<sup>68</sup>, Biyun Qian<sup>21</sup>, Chen Wu<sup>69</sup>, Daru Lu<sup>2,3</sup>, Jianjun Liu<sup>66,70,71</sup>, Ann G Schwartz<sup>72</sup>, Richard Houlston<sup>73</sup>, Margaret R Spitz<sup>74</sup>, Ivan P Gorlov<sup>74</sup>, Xifeng Wu<sup>75</sup>, Ping Yang<sup>76</sup>, Stephen Lam<sup>77</sup>, Adonina Tardon<sup>78</sup>, Chu Chen<sup>79</sup>, Stig E Bojesen<sup>80,81</sup>, Mattias Johansson<sup>82</sup>, Angela Risch<sup>83,84,85</sup>, Heike Bickeböllner<sup>86</sup>, Bu-Tian Ji<sup>4</sup>, H-Erich Wichmann<sup>87,88,89</sup>, David C. Christiani<sup>90</sup>, Gadi Rennert<sup>91</sup>, Susanne Arnold<sup>92</sup>, Paul Brennan<sup>82</sup>, James McKay<sup>82</sup>, John K Field<sup>93</sup>, Sanjay S Shete<sup>94</sup>, Loic Le Marchand<sup>95</sup>, Geoffrey Liu<sup>96</sup>, Angeline Andrew<sup>97</sup>, Lambertus A Kiemeny<sup>98</sup>, Shan Zienolddiny-Narui<sup>99</sup>, Kjell Grankvist<sup>100</sup>, Mikael Johansson<sup>101</sup>, Angela Cox<sup>102</sup>, Fiona Taylor<sup>102</sup>, Jian-Min Yuan<sup>103</sup>, Philip Lazarus<sup>104</sup>, Matthew B Schabath<sup>105</sup>, Melinda C Aldrich<sup>106</sup>, Hyo-Sung Jeon<sup>107</sup>, Shih Sheng Jiang<sup>108</sup>, Jae Sook Sung<sup>54</sup>, Chung-Hsing Chen<sup>108</sup>, Chin-Fu Hsiao<sup>7</sup>, Yoo Jin Jung<sup>109</sup>, Huan Guo<sup>110</sup>, Zhibin Hu<sup>111</sup>, Laurie Burdett<sup>4,112</sup>, Meredith Yeager<sup>4,112</sup>, Amy Hutchinson<sup>4,112</sup>, Belynda Hicks<sup>4,112</sup>, Jia Liu<sup>4,112</sup>, Bin Zhu<sup>4,112</sup>, Sonja I Berndt<sup>4</sup>, Wei Wu<sup>113</sup>, Junwen Wang<sup>114,115</sup>, Yuqing Li<sup>70</sup>, Jin Eun Choi<sup>107</sup>, Kyong Hwa Park<sup>54</sup>, Sook Whan Sung<sup>116</sup>, Li Liu<sup>117</sup>, Chang Hyun Kang<sup>109</sup>, Wen-Chang Wang<sup>118</sup>, Jun Xu<sup>119</sup>, Peng Guan<sup>113,120</sup>, Wen Tan<sup>69</sup>, Chong-Jen Yu<sup>121</sup>, Gong Yang<sup>9</sup>, Alan Dart Loon Sihoe<sup>122</sup>, Ying Chen<sup>5</sup>, Yi Young Choi<sup>107</sup>, Jun Suk Kim<sup>123</sup>, Ho-Il Yoon<sup>124</sup>, In Kyu Park<sup>109</sup>, Ping Xu<sup>125</sup>, Qincheng He<sup>113</sup>, Chih-Liang Wang<sup>126</sup>, Hsiao-Han Hung<sup>108,127</sup>, Roel C.H. Vermeulen<sup>128</sup>, Iona Cheng<sup>129</sup>, Junjie Wu<sup>2,3</sup>, Wei-Yen Lim<sup>5</sup>, Fang-Yu Tsai<sup>108,127</sup>, John K.C. Chan<sup>130</sup>, Jihua Li<sup>131</sup>, Hongyan Chen<sup>2,3</sup>, Hsien-Chih Lin<sup>7</sup>, Li Jin<sup>2,3</sup>, Jie Liu<sup>132</sup>, Norie Sawada<sup>133</sup>, Taiki Yamaji<sup>134</sup>, Kathleen Wyatt<sup>4,112</sup>, Shengchao A. Li<sup>4,112</sup>, Hongxia Ma<sup>111,135</sup>, Meng Zhu<sup>111,135</sup>, Zhehai Wang<sup>132</sup>, Sensen Cheng<sup>132</sup>, Xuelian Li<sup>113,120</sup>, Yangwu Ren<sup>113,120</sup>, Ann Chao<sup>136</sup>, Motoki Iwasaki<sup>133,134</sup>, Junjie Zhu<sup>18</sup>, Gening Jiang<sup>18</sup>, Ke Fei<sup>18</sup>, Guoping Wu<sup>53</sup>, Chih-Yi Chen<sup>137</sup>, Chien-Jen Chen<sup>138</sup>, Pan-Chyr Yang<sup>139</sup>, Jinming Yu<sup>132</sup>, Victoria L. Stevens<sup>140</sup>, Joseph F. Fraumeni Jr<sup>4</sup>

<sup>1</sup>Department of Clinical Epidemiology and Center of Evidence Based Medicine, The First Hospital of China Medical University, Shenyang, China, <sup>2</sup>Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China, <sup>3</sup>State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China, <sup>4</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA, <sup>5</sup>Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore, <sup>6</sup>Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore, <sup>7</sup>Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan, <sup>8</sup>Department of Applied Mathematics, Chung Yuan Christian University, Chong-Li, Taiwan, <sup>9</sup>Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center and Vanderbilt-Ingram Cancer Center, Nashville, TN, USA, <sup>10</sup>Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea, <sup>11</sup>Guangdong 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, <sup>12</sup>Department of Pathology, Queen Mary Hospital, Hong Kong, Hong Kong, <sup>13</sup>Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA, <sup>14</sup>Department of Biostatistics, Bioinformatics & Biomathematics, Georgetown University Medical Center,

Washington, DC, USA, <sup>15</sup>Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA, <sup>16</sup>Saville Cancer Screening and Prevention Center, Inova Schar Cancer Institute, Inova Health System, Fairfax, VA, USA, <sup>17</sup>Department of Environmental Health, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea,, <sup>18</sup>Epidemiology and Community Health Branch, National Heart Lung and Blood Institute, Bethesda, MD, USA, <sup>19</sup>Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea, <sup>20</sup>Shanghai Pulmonary Hospital, Shanghai, China, <sup>21</sup>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, <sup>22</sup>Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan, <sup>23</sup>Lung and Esophageal Cancer Clinic, Chonnam National University Hwasun Hospital, Hwasun, Republic of Korea, <sup>24</sup>Department of Internal Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea, <sup>25</sup>Department of Etiology & Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, <sup>26</sup>Department of Medicine, School of Clinical Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong, Hong Kong, <sup>27</sup>Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, <sup>28</sup>Center for Antibody and Vaccine Therapy, Research Hospital, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, <sup>29</sup>Department of Medical Oncology and Cancer Center, and Center for Advanced Medicine against Cancer, Shiga University of Medical Science, Shiga, Japan, <sup>30</sup>Department of Statistics & Research Institute of Natural Sciences, Sookmyung Women's University, Seoul, 04310, Korea, <sup>31</sup>Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, <sup>32</sup>Department of Thoracic Surgery, Tokyo Medical and Dental University, Tokyo, Japan, <sup>33</sup>Department of Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan, <sup>34</sup>Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, USA, <sup>35</sup>Department of Medical Oncology, Japanese Red Cross Medical Center, Tokyo, Japan, <sup>36</sup>Department of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan, <sup>37</sup>Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, Yokohama, Japan, <sup>38</sup>Department of Thoracic Surgery, Kanagawa Cancer Center, Yokohama, Japan, <sup>39</sup>Department of Thoracic Oncology, National Cancer Center Hospital East, Kashiwa, Japan, <sup>40</sup>Department of Thoracic Surgery, National Cancer Center Hospital East, Kashiwa, Japan, <sup>41</sup>Division of Medical AI Research and Development, National Cancer Center Research Institute, Tokyo, Japan, <sup>42</sup>Department of Cellular and Organ Pathology, Graduate School of Medicine, Akita University, Akita, Japan, <sup>43</sup>Department of Thoracic Surgery, Graduate School of Medicine, Akita University, Akita, Japan, <sup>44</sup>Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, Japan, <sup>45</sup>Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan, <sup>46</sup>Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Science, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, <sup>47</sup>Department of Molecular Biology, Institute for Advanced Medical Sciences, Nippon Medical School, Tokyo, Japan, <sup>48</sup>Department of Surgery, Division of General Thoracic Surgery, Shinshu University School of Medicine Asahi, Nagano, Japan, <sup>49</sup>Department of Chest Surgery, Fukushima Medical University School of Medicine, Fukushima, Japan, <sup>50</sup>Department of Gastrointestinal Tract Surgery, Fukushima Medical University School of

Medicine, Fukushima, Japan, <sup>51</sup>Department of Integrative center of General Surgery, Gunma University Hospital, Gunma, Japan, <sup>52</sup>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, <sup>53</sup>China National Environmental Monitoring Center, Beijing, China, <sup>54</sup>Department of Internal Medicine, Division of Oncology/Hematology, College of Medicine, Korea University Anam Hospital, Seoul, Republic of Korea, <sup>55</sup>Department of Clinical Oncology, The University of Hong Kong, Queen Mary Hospital, Hong Kong, Hong Kong, <sup>56</sup>School of Medicine and Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan, <sup>57</sup>Department of Internal Medicine, Division of Pulmonary Medicine, Chung Shan Medical University Hospital, Taichung, Taiwan, <sup>58</sup>Institute of Biomedical Sciences, National Chung Hsing University, Taichung, Taiwan, <sup>59</sup>Department of Internal Medicine, Division of Chest Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, <sup>60</sup>Department of Respiratory Therapy, Chang Gung University, Taoyuan, Taiwan, <sup>61</sup>Department of Pulmonary and Critical Care, Xiamen Chang Gung Hospital, Xiamen, China, <sup>62</sup>Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan, <sup>63</sup>Department of Internal Medicine, E-Da Cancer Hospital, I-Shou University and Kaohsiung Medical University, Kaohsiung, Taiwan, <sup>64</sup>Department of Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan, <sup>65</sup>Department of Chest Medicine, Taipei Veterans General Hospital, and school of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, <sup>66</sup>Lung Cancer Center, Kyungpook National University Medical Center, Daegu, Republic of Korea, <sup>67</sup>Jeonnam Regional Cancer Center, Chonnam National University, Hwasun, Republic of Korea, <sup>68</sup>Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China, <sup>69</sup>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, <sup>70</sup>Department of Human Genetics, Genome Institute of Singapore, Singapore, Singapore, <sup>71</sup>School of Life Sciences, Anhui Medical University, Hefei, China, <sup>72</sup>Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA, <sup>73</sup>Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK, <sup>74</sup>Department of Medicine, Section of Epidemiology and Population Science, Institute for Clinical and Translational Research, Houston, TX, USA, <sup>75</sup>School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, <sup>76</sup>Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, USA, <sup>77</sup>Department of Integrative Oncology, BC Cancer, Vancouver, BC, Canada, <sup>78</sup>IUOPA, University of Oviedo and CIBERESP, Spain, <sup>79</sup>Program in Epidemiology, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA, <sup>80</sup>Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, <sup>81</sup>Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark, <sup>82</sup>International Agency for Research on Cancer (IARC/WHO), Lyon, France, <sup>83</sup>German Cancer Research Center (DKFZ), Heidelberg, Germany, <sup>84</sup>Translational Lung Research Center Heidelberg (TLRC-H), Member of the German Center for Lung Research (DZL), Heidelberg, Germany, <sup>85</sup>University of Salzburg and Cancer Cluster Salzburg, Salzburg, Austria, <sup>86</sup>University Medical Center Goettingen, Goettingen, Germany, <sup>87</sup>Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany, <sup>88</sup>Helmholtz Center Munich, Institute of Epidemiology, Munich, Germany, <sup>89</sup>Institute of Medical Statistics and Epidemiology, Technical University Munich, Munich, Germany, <sup>90</sup>Harvard TH Chan School of Public Health, Boston,

Massachusetts, USA, <sup>91</sup>Carmel Medical Center, Israel, <sup>92</sup>Markey Cancer Center, Lexington, KY, USA, <sup>93</sup>Liverpool University, Liverpool, UK, <sup>94</sup>The University of Texas MD Anderson Cancer Center, Houston, Texas, USA, <sup>95</sup>Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA, <sup>96</sup>Princess Margaret Cancer Center, Toronto, ON, Canada, <sup>97</sup>Norris Cotton Cancer Center, Lebanon, NH, USA, <sup>98</sup>Radboud University Medical Center, Nijmegen, Netherlands, <sup>99</sup>National Institute of Occupational Health, Oslo, Norway, <sup>100</sup>Department of Medical Biosciences, Umeå University, Umeå, Sweden, <sup>101</sup>Department of Radiation Sciences, Umeå University, Umeå, Sweden, <sup>102</sup>University of Sheffield, Sheffield, UK, <sup>103</sup>UPMC Hillman Cancer Center and Department of Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA, <sup>104</sup>Department of Pharmaceutical Sciences, College of Pharmacy and Pharmaceutical Sciences, Washington State University, Spokane, WA, USA, <sup>105</sup>Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA, <sup>106</sup>Department of Thoracic Surgery, Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA, <sup>107</sup>Cancer Research Center, Kyungpook National University Medical Center, Daegu, Republic of Korea, <sup>108</sup>National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan, <sup>109</sup>Department of Thoracic and Cardiovascular Surgery, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea, <sup>110</sup>Department of Occupational and Environmental Health and Ministry of Education Key Lab for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, <sup>111</sup>Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China, <sup>112</sup>Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc, Rockville, MD, USA, <sup>113</sup>Department of Epidemiology, School of Public Health, China Medical University, Shenyang, China, <sup>114</sup>Division of Applied Oral Sciences & Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China, <sup>115</sup>State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, China, <sup>116</sup>State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, China, <sup>117</sup>Department of Oncology, Cancer Center, Union Hospital, Huazhong University of Science and Technology, Wuhan, China, <sup>118</sup>The Ph.D. Program for Translational Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan, <sup>119</sup>School of Public Health, Li Ka Shing (LKS) Faculty of Medicine, The University of Hong Kong, Hong Kong, China, <sup>120</sup>Key Laboratory of Cancer Etiology and Intervention, University of Liaoning Province, Shenyang, China, <sup>121</sup>Department of Internal Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan, <sup>122</sup>Gleneagles Hong Kong Hospital, Hong Kong, China, <sup>123</sup>Department of Internal Medicine, Division of Medical Oncology, College of Medicine, Korea University Guro Hospital, Seoul, Republic of Korea, <sup>124</sup>Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea, <sup>125</sup>Department of Oncology, Wuhan Iron and Steel (Group) Corporation Staff-Worker Hospital, Wuhan, China, <sup>126</sup>Division of Pulmonary Oncology and Interventional Bronchoscopy, Department of Thoracic Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan, <sup>127</sup>Division of Thoracic Surgery, Department of Surgery, Chung Shan Medical University Hospital, Taichung, Taiwan, <sup>128</sup>Division of Environmental Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands, <sup>129</sup>Department of Epidemiology and Biostatistics, University of

California, San Francisco, San Francisco, CA, USA, <sup>130</sup>Department of Pathology, Queen Elizabeth Hospital, Hong Kong, China, <sup>131</sup>Qijing Center for Diseases Control and Prevention, Qijing, China, <sup>132</sup>Department of Oncology, Shandong Cancer Hospital and Institute, Shandong Academy of Medical Sciences, Jinan, China, <sup>133</sup>Division of Cohort Research, National Cancer Center Institute for Cancer Control, Tokyo, Japan, <sup>134</sup>Division of Epidemiology, National Cancer Center Institute for Cancer Control, Tokyo, Japan, <sup>135</sup>Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China, <sup>136</sup>Center for Global Health, National Cancer Institute, Bethesda, MD, USA, <sup>137</sup>Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan, <sup>138</sup>Genomic Research Center, Academia Sinica, Taipei, Taiwan, <sup>139</sup>Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan, <sup>140</sup>Laboratory Services, American Cancer Society, Georgia, USA