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

Batel Blechter PhD [†,1], Xiaoyu Wang PhD [†,1], Juncheng Dai PhD [†,2,3], Christiana Karsonaki DPhil [†,4], Jianxin Shi PhD [†,1], Kouya Shiraishi PhD [†,5], Jiyeon Choi PhD [†,1], Keitaro Matsuo PhD [†,6], Tzu-Yu Chen PhD [†,7,8], Rayjean J Hung PhD, MS [†,9], Kexin Chen PhD [†,10], Xiao-Ou Shu MD, PhD [†,11], Young Tae Kim MD, PhD[12], Parichoy Pal Choudhury PhD[13], Jacob Williams PhD[1], Maria Teresa Landi MD, PhD[1], Dongxin Lin PhD[14], Wei Zheng MD, PhD[11], Zhihua Yin PhD[15], Baosen Zhou PhD[16], Jiucun Wang PhD[17,18], Wei Jie Seow PhD[19,20], Lei Song PhD[1], I-Shou Chang PhD[21], Wei Hu PhD[1], Li-Hsin Chien PhD[7,22], Qiuyin Cai PhD[11], Yun-Chul Hong PhD[23], Hee Nam Kim PhD[24], Yi-Long Wu PhD[25], Maria Pik Wong PhD[26], Brian Douglas Richardson PhD[1,27], Shilan Li PhD[1,28], Tongwu Zhang PhD[1], Charles Breeze PhD[1], Zhaoming Wang PhD, MSc[29], Bryan A Bassig PhD[30], Jin Hee Kim PhD[31], Demetrius Albanes MD[1], Jason YY Wong SM, ScD[32], Min-Ho Shin PhD[24], Lap Ping Chung PhD[26], Yang Yang PhD[33], Hong Zheng PhD[10], Hongji Dai PhD[10], Yasushi Yatabe PhD[34], Xu-Chao Zhang MD, PhD[25], Young-Chul Kim MD, PhD[35,36], Neil E Caporaso MD[1], Jiang Chang PhD[37], James Chung Man Ho MD[38], Yataro Daigo MD, PhD[39,40], Yukihide Momozawa PhD[41], Yoichiro Kamatani MD, PhD[42], Masashi Kobayashi MD, PhD[43], Kenichi Okubo MD, PhD[43], Takayuki Honda MD, PhD[44], H Dean Hosgood PhD[45], Hideo Kunitoh MD[46], Shun-ichi Watanabe MD[47], Yohei Miyagi MD, PhD[48], Shingo Matsumoto PhD[49], Hidehito Horinouchi PhD[47], Masahiro Tsuboi MD, PhD[50], Ryuji Hamamoto PhD[51], Koichi Goto MD, PhD[49], Atsushi Takahashi PhD[52], Akiteru Goto MD, PhD[53], Yoshihiro Minamiya PhD[54], Megumi Hara PhD[55], Yuichiro Nishida PhD[55], Kenji Takeuchi PhD[56], Kenji Wakai PhD[56], Koichi Matsuda PhD[57], Yoshinori Murakami MD, PhD[58], Kimihiro Shimizu PhD[59], Hiroyuki Suzuki MD, PhD[60], Motonobu Saito MD, PhD[61], Yoichi Ohtaki PhD[62], Kazumi Tanaka MD, PhD[63], Tangchun Wu MD, PhD[64], Fusheng Wei MD[65], Mitchell J Machiela ScD[1], Yeul Hong Kim MD, PhD[66], In-Jae Oh MD, PhD[67,36], Victor Ho Fun Lee PhD[68], Gee-Chen Chang PhD[69,70,71,72,73], Kuan-Yu Chen MD, PhD[74], Wu-Chou Su MD[75], Yuh-Min Chen MD, PhD[76], Adeline Seow MBBS, MD[19], Jae Yong Park MD, PhD[77], Sun-Seog Kweon PhD[24,78], Yu-Tang Gao MD[79], Jianjun Liu PhD[80,81], Ann G Schwartz PhD, MPH[82], Richard Houlston PhD[83], Ivan P Gorlov PhD[84], Xifeng Wu PhD[85], Ping Yang MD, PhD, MPH[86], Stephen Lam PhD[87], Adonina Tardon PhD[88], Chu Chen PhD[89], Stig E Bojesen PhD[90,91], Mattias Johansson PhD[92], Angela Risch PhD[93,94,95], Heike Bickeböller PhD[96], Bu-Tian Ji PhD[1], H-Erich Wichmann PhD[97,98], David C Christiani MD, MPH, MS[99], Gad Rennert MD, PhD[100], Susanne M Arnold MD[101], Paul Brennan PhD[92], James McKay PhD[92], John K Field PhD[102], Michael P.A. Davies PhD[102], Sanjay S Shete PhD[103], Loïc Le Marchand MD, PhD[104], Geoffrey Liu PhD[105], Angeline S Andrew PhD[106], Lambertus A Kiemeney PhD[107], Shan Zienolddiny-Narui PhD[108], Kjell Grankvist PhD[109], Angela Cox PhD[110], Fiona Taylor PhD[111], Jian-Min Yuan PhD[112], Philip Lazarus PhD[113], Matthew B Schabath PhD[114], Melinda C Aldrich MPH, PhD, FACE[115], Hyo-Sung Jeon PhD[116], Shih Sheng Jiang PhD[21], Chung-Hsing Chen PhD[21,117], Chin-Fu Hsiao PhD[7], Zhibin Hu PhD[2], Laura Burdett PhD[1,118], Meredith Yeager PhD[1,118], Amy Hutchinson PhD[1,118], Belynda Hicks PhD[1,118], Jia Liu PhD[1,118], Sonja I Berndt PhD[1], Wei Wu PhD[15], Junwen Wang PhD[119,120], Yuqing Li PhD[121], Jin Eun Choi PhD[122], Kyong Hwa Park MD, PhD[123], Sook Whan Sung MD[124], Chang Hyun Kang PhD[125], Wen-Chang Wang PhD[126], Jun Xu PhD[127], Peng Guan PhD[15,128], Wen Tan PhD[14], Chong-Jen Yu MD, PhD[129], Gong Yang MD[11], Alan Dart Loon Sihoe PhD[130], Yi Young Choi PhD[131], In Kyu Park MD, PhD[125], Hsiao-Han Hung MS[21], Roel C.H.

46 Vermeulen PhD[132], Iona Cheng PhD[121], Junjie Wu MD[133], Fang-Yu Tsai MS[21], John K.C. Chan
47 MD[134], Jihua Li MD[135,136], Hsien-Chih Lin PhD[7,136], Jie Liu PhD[137], Bao Song PhD[137], Norie
48 Sawada MD, PhD[138], Taiki Yamaji MD, PhD[139], Kathleen Wyatt MS[1,118], Hongxia Ma PhD[2,3],
49 Meng Zhu PhD[2,3], Yifan Wang PhD[2,3], Tianchen Qi MS[2,3], Xuelian Li PhD[15,128], Yangwu Ren
50 PhD[15,128], Ann Chao PhD[140], Motoki Iwasaki MD, PhD[138,139], Junjie Zhu PhD[33], Guoping Wu
51 MD[65], Chih-Yi Chen MD[141,142], Chien-Jen Chen ScD, PhD[143], Pan-Chyr Yang MD, PhD[144],
52 Victoria L. Stevens PhD[145], Joseph F. Fraumeni Jr PhD[1], Kuang Lin PhD[4], Robin G Walters
53 PhD[4], Zhengming Chen DPhil[4], Nilanjan Chatterjee PhD [‡,1,146,147], Olga Y Gorlova PhD
54 [‡,148,149,150], Christopher I Amos PhD [‡,148,149,150,151], Hongbing Shen PhD [‡,2,3], Chao Agnes Hsiung
55 PhD [‡,7], Stephen J Chanock MD [‡,1], Nathaniel Rothman MD, MPH, MHS [‡,1], Takashi Kohno
56 PhD [‡,5], Qing Lan MD, PhD [‡,1], Haoyu Zhang PhD [‡,1]
57
58 [1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD,
59 USA, [2]Department of Epidemiology, School of Public Health, Nanjing Medical University,
60 Nanjing, China, [3]Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment,
61 Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing,
62 China, [4]Nuffield Department of Population Health, University of Oxford, Oxford, UK, [5]Division
63 of Genome Biology, National Cancer Research Institute, Tokyo, Japan, [6]Division of Molecular
64 Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan, [7]Institute of Population
65 Health Sciences, National Health Research Institutes, Zhunan, Taiwan, [8]Department of Smart
66 Computing and Applied Mathematics, Tunghai University, Taichung, Taiwan, [9]Prosserman
67 Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health,
68 Toronto, Canada, [10]Department of Epidemiology and Biostatistics, Key Laboratory of Prevention
69 and Control of Human Major Diseases of the Ministry of Education, Key Laboratory of
70 Molecular Cancer Epidemiology of Tianjin, National Clinical Research Center for Cancer,
71 Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin,
72 China, [11]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical
73 Center and Vanderbilt-Ingram Cancer Center, Nashville, TN, USA, [12]Department of Thoracic
74 and Cardiovascular Surgery, Seoul National University College of Medicine, Seoul National
75 University Hospital, Seoul, Republic of Korea, [13]American Cancer Society, Atlanta, GA, USA,
76 [14]Department of Etiology & Carcinogenesis and State Key Laboratory of Molecular Oncology,
77 Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical
78 College, Beijing, China, [15]Department of Epidemiology, School of Public Health, China Medical
79 University, Shenyang, China, [16]Department of Clinical Epidemiology and Center of Evidence
80 Based Medicine, The First Hospital of China Medical University, Shenyang, China, [17]Ministry of
81 Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan
82 University, Shanghai, China, [18]State Key Laboratory of Genetics and Development of Complex
83 Phenotypes, School of Life Sciences, Fudan University, Shanghai, China, [19]Saw Swee Hock
84 School of Public Health, National University of Singapore and National University Health
85 System, Singapore, Singapore, [20]Department of Medicine, Yong Loo Lin School of Medicine,
86 National University of Singapore and National University Health System, Singapore, Singapore,
87 [21]National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan,
88 [22]Department of Applied Mathematics, National Dong Hwa University, Hualien, Taiwan,
89 [23]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul,
90 Republic of Korea, [24]Department of Preventive Medicine, Chonnam National University Medical
91 School, Gwangju, Republic of Korea, [25]Guangdong Lung Cancer Institute, Guangdong

Provincial People's Hospital & Guangdong Academy of Medical Sciences, Southern Medical University, Guangzhou, China, [26]Department of Pathology, Queen Mary Hospital, Hong Kong, China, [27]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA, [28]Department of Biostatistics, Bioinformatics & Biomathematics, Georgetown University Medical Center, Washington DC, DC, USA, [29]Department of Epidemiology and Cancer Control (Primary) & Department of Computational Biology (Secondary), St. Jude Children's Research Hospital, Memphis, TN, USA, [30]Saville Cancer Screening and Prevention Center, Inova Schar Cancer Institute, Inova Health System, Fairfax, VA, USA, [31]Department of Integrative Bioscience & Biotechnology, Institute of Bioscience, Sejong University, Seoul, 05006, Republic of Korea, [32]Epidemiology and Community Health Branch, National Heart Lung and Blood Institute, Bethesda, MD, USA, [33]Shanghai Pulmonary Hospital, Shanghai, China, [34]Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan, [35]Lung Cancer Clinic, Pulmonary Medicine, Chonnam National University Medical School, Hwasun Hospital, Hwasun, Republic of Korea, [36]Department of Internal Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea, [37]School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [38]Department of Medicine, School of Clinical Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong, China, [39]Center for Antibody and Vaccine Therapy, Research Hospital, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [40]Department of Medical Oncology and Cancer Center, and Center for Advanced Medicine against Cancer, Shiga University of Medical Science, Shiga, Japan, [41]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [42]Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, [43]Department of Thoracic Surgery, Tokyo Medical and Dental University, Tokyo, Japan, [44]Department of Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan, [45]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, USA, [46]Department of Medical Oncology, Japanese Red Cross Medical Center, Tokyo, Japan, [47]Department of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan, [48]Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, Yokohama, Japan, [49]Department of Thoracic Oncology, National Cancer Center Hospital East, Kashiwa, Japan, [50]Department of Thoracic Oncology, Division of Thoracic Surgery, National Cancer Center Hospital East, Kashiwa, Japan, [51]Division of Medical AI Research and Development, National Cancer Center Research Institute, Tokyo, Japan, [52]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [53]Department of Cellular and Organ Pathology, Graduate School of Medicine, Akita University, Akita, Japan, [54]Department of Thoracic Surgery, Graduate School of Medicine, Akita University, Akita, Japan, [55]Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, Japan, [56]Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan, [57]Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Science, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, [58]Department of Molecular Biology, Institute for Advanced Medical Sciences, Nippon Medical School, Tokyo, Japan, [59]Division of General Thoracic Surgery, Department of Surgery, Shinshu University School of Medicine Asahi, Nagano, Japan, [60]Department of Chest Surgery, Fukushima Medical University, Fukushima, Japan, [61]Department of Cancer Genome Medicine, Fukushima Medical University Hospital, Fukushima, Japan,

138 [62]Department of Integrative center of General Surgery, Gunma University Hospital, Gunma,
139 Japan, [63]Department of Healthcare Quality and Safety, Gunma University Graduate School of
140 Medicine, Gunma, Japan, [64]Dept of Occupational and Environmental Health, Key Laboratory for
141 Environment and Health, Ministry of Education and State Key Laboratory of Environmental
142 Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan,
143 China, [65]China National Environmental Monitoring Center, Beijing, China, [66]Korea University
144 College of Medicine Anam Hospital, Seoul, Republic of Korea, [67]Lung Cancer Center, Chonnam
145 National University Hwasun Hospital, Hwasuneup, Republic of Korea, [68]Department of Clinical
146 Oncology, The University of Hong Kong, Queen Mary Hospital, Hong Kong, China, [69]School of
147 Medicine and Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan, [70]Lung
148 Cancer Diagnosis & Treatment Research Center, Chung Shan Medical University Hospital,
149 Taichung, Taiwan, [71]Institute of Biomedical Sciences, National Chung Hsing University,
150 Taichung, Taiwan, [72]Department of Internal Medicine, Division of Chest Medicine, Taichung
151 Veterans General Hospital, Taichung, Taiwan, [73]Department of Pulmonary and Critical Care,
152 Xiamen Chang Gung Hospital, Xiamen, China, [74]Department of Internal Medicine, National
153 Taiwan University Hospital and College of Medicine, Taipei, Taiwan, [75]Department of
154 Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng
155 Kung University, Tainan, Taiwan, [76]Department of Chest Medicine, Taipei Veterans General
156 Hospital, and School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan,
157 [77]Lung Cancer Center, Kyungpook National University Medical Center, Daegu, Republic of
158 Korea, [78]Jeonnam Regional Cancer Center, Chonnam National University, Hwasun, Republic of
159 Korea, [79]Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai
160 Jiaotong University, School of Medicine, Shanghai, China, [80]Lab of Human Genomics, Genome
161 Institute of Singapore, A*STAR, Singapore, Singapore, [81]Yong Loo Lin School of Medicine,
162 National University of Singapore, Singapore, Singapore, Singapore, [82]Karmanos Cancer
163 Institute, Wayne State University, Detroit, MI, USA, [83]Division of Genetics and Epidemiology,
164 Institute of Cancer Research, London, UK, [84]Department of Medicine, Section of Epidemiology
165 and Population Science, Institute for Clinical and Translational Research, Houston, TX, USA,
166 [85]School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China, [86]Department of
167 Quantitative Health Sciences, Mayo Clinic, Scottsdale, AZ, USA, [87]Department of Integrative
168 Oncology, BC Cancer, Vancouver, BC, Canada, [88]Health Research Institute of Asturias (ISPA),
169 IUOPA University of Oviedo and University Nebrija, Spain, [89]Program in Epidemiology, Public
170 Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA, [90]Department Of
171 Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen,
172 Copenhagen, Denmark, [91]Department of Clinical Biochemistry, Herlev and Gentofte Hospital,
173 Copenhagen University Hospital, Copenhagen, Denmark, [92]International Agency for Research
174 on Cancer (IARC/WHO), Lyon, France, [93]German Cancer Research Center (DKFZ), Heidelberg,
175 Germany, [94]Translational Lung Research Center Heidelberg (TLRC-H), Member of the German
176 Center for Lung Research (DZL), Heidelberg, Germany, [95]Department of Biosciences and
177 Medical Biology and Center for Tumorbiology and Immunology (CTBI), University of Salzburg
178 and Cancer Cluster Salzburg, Salzburg, Austria, [96]University Medical Center Goettingen,
179 Goettingen, Germany, [97]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig
180 Maximilians University, Munich, Germany, [98]Helmholtz Center Munich, Institute of
181 Epidemiology, Munich, Germany, [99]Harvard TH Chan School of Public Health, Boston, MA,
182 USA, [100]Technion-Israel Institute of Technology and APRPM, Haifa, Israel, [101]Markey Cancer
183 Center, Lexington, KY, USA, [102]University of Liverpool, Liverpool, UK, [103]The University of

184 Texas, MD Anderson Cancer Center, Houston, TX, USA, [104]Epidemiology Program, University
185 of Hawaii Cancer Center, Honolulu, HI, USA, [105]Princess Margaret Cancer Center, Toronto, ON,
186 Canada, [106]Norris Cotton Cancer Center, Lebanon, NH, USA, [107]Radboud University Medical
187 Center, Nijmegen, The Netherlands, [108]National Institute of Occupational Health, Oslo, Norway,
188 [109]Department of Medical Biosciences, Umeå University, Umeå, Sweden, [110]Division of Clinical
189 Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield, UK,
190 [111]University of Sheffield, Sheffield, UK, [112]UPMC Hillman Cancer Center and Department of
191 Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA,
192 [113]Department of Pharmaceutical Sciences, School of Pharmacy and Pharmaceutical Sciences,
193 SUNY University at Buffalo, NY, USA, [114]Department of Cancer Epidemiology, H. Lee Moffitt
194 Cancer Center and Research Institute, Tampa, FL, USA, [115]Department of Medicine, Division of
195 Genetic Medicine and Clinical Pharmacology, Vanderbilt University Medical Center, Nashville,
196 TN, USA, [116]Cancer Research Center, Kyungpook National University Medical Center, Daegu,
197 Republic of Korea, [117]Department of Mathematics, University of Taipei, Taipei, Taiwan,
198 [118]Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research,
199 Leidos Biomedical Research Inc, Rockville, MD, USA, [119]Division of Applied Oral Sciences &
200 Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong,
201 China, [120]State Key LaState Key Laboratory of Pharmaceutical Biotechnology, The University of
202 Hong Kong, Hong Kong, China, [121]Department of Epidemiology and Biostatistics, University of
203 California, San Francisco, San Francisco, CA, USA, [122]Department of Biochemistry and Cell
204 Biology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea,
205 [123]Department of Internal Medicine, Division of Oncology/Hematology, College of Medicine,
206 Korea University Anam Hospital, Seoul, Republic of Korea, [124]Department of Thoracic and
207 Cardiovascular Surgery, Seoul National University Bundang Hospital, Seongnam, Republic of
208 Korea, [125]Department of Thoracic and Cardiovascular Surgery, Cancer Research Institute, Seoul
209 National University College of Medicine, Seoul, Republic of Korea, [126]The Ph.D. Program for
210 Translational Medicine, College of Medical Science and Technology, Taipei Medical University,
211 Taipei, Taiwan, [127]School of Public Health, Li Ka Shing (LKS) Faculty of Medicine, The
212 University of Hong Kong, Hong Kong, China, [128]Key Laboratory of Cancer Etiology and
213 Intervention, University of Liaoning Province, Shenyang, China, [129]Department of Internal
214 Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan,
215 [130]Gleneagles Hong Kong Hospital, Hong Kong, China, [131]Cancer Research Center, Kyungpook
216 National University Medical Center, Daegu, Republic of Korea, [132]Division of Environmental
217 Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The
218 Netherlands, [133]Department of Pulmonary and Critical Care Medicine Zhongshan Hospital,
219 Fudan University Shanghai Geriatric Medical Center, Shanghai, China, [134]Department of
220 Pathology, Queen Elizabeth Hospital, Hong Kong, China, [135]Qujing Center for Diseases Control
221 and Prevention, Qujing, China, [136]State Key Laboratory of Genetic Engineering, School of Life
222 Sciences, Fudan University, Shanghai, China, [137]Department of Oncology, Shandong Cancer
223 Hospital and Institute, Shandong Academy of Medical Sciences, Jinan, China, [138]Division of
224 Cohort Research, National Cancer Center Institute for Cancer Control, Tokyo, Japan, [139]Division
225 of Epidemiology, National Cancer Center Institute for Cancer Control, Tokyo, Japan, [140]Center
226 for Global Health, National Cancer Institute, Bethesda, MD, USA, [141]Institute of Medicine,
227 Chung Shan Medical University, Taichung, Taiwan, [142]Department of Surgery, Division of
228 Thoracic Surgery, Chung Shan Medical University Hospital, Taichung, Taiwan, [143]Genomic
229 Research Center, Academia Sinica, Taipei, Taiwan, [144]Department of Internal Medicine,

230 National Taiwan University Hospital, Taipei, Taiwan, [145]Laboratory Services, American Cancer
231 Society, Georgia, USA, [146]Department of Oncology, School of Medicine, Johns Hopkins
232 University, Baltimore, MD, USA, [147]Department of Biostatistics, Johns Hopkins Bloomberg
233 School of Public Health, Baltimore, MD, USA, [148]Institute for Clinical and Translational
234 Research, Baylor College of Medicine, Houston, TX, USA, [149]Section of Epidemiology and
235 Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA,
236 [150]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX,
237 USA, [151]University of New Mexico, Comprehensive Cancer Center, Albuquerque, NM, USA
238

239 † These authors contributed equally to this work

240 ‡ These authors jointly directed this work

241

242 Correspondence to:

243 Batel Blechter

244 National Cancer Institute

245 9609 Medical Center Drive, Room 6E124

246 Rockville MD, 20850, USA

247 Email: batel.blechter@nih.gov
248
249
250
251

Abstract

Background

Lung adenocarcinoma (LUAD) in never-smokers is a major public health burden, especially among East Asian women. Polygenic risk scores (PRSs) are promising for risk stratification but are primarily developed in European-ancestry populations. We aimed to develop and validate single- and multi-ancestry PRSs for East Asian never-smokers to improve LUAD risk prediction.

Methods

PRSs were developed using genome-wide association study summary statistics from East Asian (8,002 cases; 20,782 controls) and European (2,058 cases; 5,575 controls) populations. Single-ancestry models included PRS-25, PRS-CT, and LDpred2; multi-ancestry models included LDpred2+PRS-EUR128, PRS-CSx, and CT-SLEB. Performance was evaluated in independent East Asian data from the Female Lung Cancer Consortium (FLCCA) and externally validated in the Nanjing Lung Cancer Cohort (NJLCC). We assessed predictive accuracy via AUC, with 10-year and (age 30-80) absolute risks estimates.

Results

The best multi-ancestry PRS, using East Asian and European data via CT-SLEB (clumping and thresholding, super learning, empirical Bayes), outperformed the best East Asian-only PRS (LDpred2; AUC=0.629, 95% CI:0.618,0.641), achieving an AUC of 0.640 (95% CI:0.629,0.653) and odds ratio of 1.71 (95% CI:1.61,1.82) per SD increase. NJLCC Validation confirmed robust performance (AUC =0.649, 95% CI: 0.623, 0.676). The top 20% PRS group had a 3.92-fold higher LUAD risk than the bottom 20%. Further, the top 5% PRS group reached a 6.69%

lifetime absolute risk. Notably, this group reached the average population 10-year LUAD risk at age 50 (0.42%) by age 41, nine years earlier.

Conclusions

Multi-ancestry PRS approaches enhance LUAD risk stratification in East Asian never-smokers, with consistent external validation, suggesting future clinical utility.

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

Abbreviations:

LUAD, lung adenocarcinoma; GWAS, genome-wide association study; EUR, European; PRS, polygenic risk score; SNP, single nucleotide polymorphism; FLCCA, Female Lung Cancer Consortium in Asia; LD, linkage disequilibrium; AUC, area under the receiver operating curve; PC, principal component; OR, odds ratio; CI, confidence interval; USPSTF, United States Preventive Services Taskforce

297 **Introduction**
298

299 Lung cancer is a major global health challenge, responsible for about 20% of all cancer
300 deaths in 2020[1]. While smoking is the primary etiologic factor, around 25% of lung cancer cases
301 occur in never-smokers, with significant geographical variations[2] . Notably, in East Asia, never-
302 smoking women exhibit high incidences of lung adenocarcinoma (LUAD), the predominant
303 histologic subtype[3].
304 Beyond smoking, risk factors for LUAD include environmental and occupational
305 exposures, lifestyle, family history, and genetic susceptibility[4–7]. Genome-wide association
306 studies (GWAS) for never-smoking lung cancer across East Asian and European (EUR)
307 populations have identified 28 susceptibility loci at 25 independent regions[8–13]. These GWAS
308 findings enhance our understanding of LUAD's genetic architecture; yet, the translation of these
309 findings into clinical applications requires further investigation.
310 Polygenic risk scores (PRS) aggregate the effects of individual common single nucleotide
311 polymorphisms (SNPs) to estimate genetic risk for diseases, potentially playing a role in cancer
312 prevention and screening. Previous efforts primarily utilized European populations for lung
313 cancer PRS development[14,15] to better identify high-risk individuals. Nonetheless, because of
314 differences in linkage disequilibrium and allele frequencies across ancestries, such PRSs often
315 underperform in East Asian never-smokers, reflecting the need to address the disproportionate
316 burden of LUAD in this population and to advance the integration of PRS into risk stratification
317 for lung cancer prevention and screening. Despite ongoing efforts to integrate PRS into clinical
318 practices for complex diseases, as seen in initiatives like the electronic Medical Records and
319 Genomics (eMERGE) network[16], Veterans Affairs Genomic Medicine at Veteran Affairs
320 (GenoVA) study[17], and the Women Informed to Screen Depending On Measures of risk

321 (WISDOM) study[18], lung cancer has been excluded. The predominant global burden of lung
322 cancer, driven primarily by smoking, overlooks the significant impact and potential benefits of
323 PRS in never-smoking populations.

324 The recent expansion of GWAS across diverse populations, combined with
325 advancements in PRS methodology, underscores a pivotal shift towards enhancing disease
326 outcome prediction beyond European populations[19]. Notably, the development of multi-ancestry
327 PRS methods marks a considerable advancement, leveraging data from various ancestral
328 backgrounds to enrich the predictive accuracy and robustness of PRSs[20–26]. By jointly modeling
329 genetic data from multiple populations, these approaches improve the predictive power of PRSs,
330 thereby enhancing disease outcome predictions in non-European populations.

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

337 This study aims to evaluate the predictive performance of single- and multi-ancestry
338 PRSs for LUAD in never-smoking East Asian individuals using state-of-the-art methodologies.
339 By integrating multiple datasets, we constructed PRSs based on summary statistics for 5,622
340 never-smoking cases and 21,813 never-smoking controls from East Asian and European
341 ancestries. We assessed the performance of these PRSs using independent, individual level data
342 of 4,438 never-smoking East Asian cases and 4,544 never-smoking East Asian controls.
343 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 East Asian 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 have been previously described in detail[27]. Briefly, biological sex and main continental ancestry were genetically determined for all participants in each study. East Asian single-ancestry PRSs for LUAD were constructed using summary data from 3,564 never-smoking LUAD cases and 16,238 never-smoking controls of East Asian ancestry from the Nanjing Lung Cancer Cohort (NJLCC)[28–31], National Cancer Center of JAPAN (NCC), and the Research Institute and Aichi Cancer Center (ACC). The NJLCC study combined data from several cities, genotyped by Affymetrix Genome-Wide Human SNP Array 6.0, Illumina OncoArray, Illumina Asian Screening Array, as well as whole genome-sequecing by Illumina HiSeq or NovaSeq platforms[28–31].The NCC study included lung cancer patients and control data from multiple Japanese studies, with genotyping conducted on Illumina HumanOmniExpress and HumanOmni1-Quad genotyping platforms. Similarly, the ACC study pooled data from multiple Japanese medical institutions and the Nagahama Study, with genotyping performed using Illumina 610k and Illumina660k platforms. Imputation and quality control were conducted using

367 the pipeline described in Shi et al. (2023)[27]. SNPs with minor allele frequency < 0.01, Hardy-
368 Weinberg Equilibrium p-value 1.0×10^{-6} , or imputation quality scores < 0.5 were excluded.
369 Genotype imputation was performed using IMPUTE2 and the 1000 Genomes Project East Asian
370 samples (Phase 3) as the reference panel. Variants present in the training datasets (NJLCC, NCC,
371 and ACC) but not in the validation dataset (FLCCA) were excluded from analyses. To construct
372 the multi-ancestry PRS, we further incorporated GWAS summary statistics from 2,058 never-
373 smoking LUAD cases and 5,575 never-smoking controls of European ancestry, genotyped using
374 Illumina Infinium OmniExpress-24 v1.2 BeadChips and Illumina Human660W-Quad
375 BeadChip[13]. Weights for all East Asian single-ancestry PRSs and the East Asian portion of the
376 multi-ancestry PRSs were estimated from a meta-analysis restricted to never-smokers from
377 NJLCC, NCC and ACC studies.

378

379 Tuning and validation: study population and genotyping

380 For PRS tuning and internal validation, we used 4,438 never-smoking LUAD cases and
381 4,544 never-smoking controls from the Female Lung Cancer Consortium in Asia (FLCCA)[10,11],
382 an international consortium composed of never-smoking East Asian women from regions
383 including Mainland China, Singapore, Taiwan, Hong Kong, South Korea and Japan [10,11]. All
384 samples were genotyped using the Illumina 660W, 370K or 610Q microarrays. Biological sex
385 and ancestry were genetically determined, and information on age and family history of lung
386 cancer were collected through a questionnaire. The FLCCA data were randomly and equally
387 divided for tuning (2,219 cases, 2,272 controls) and validation (2,219 cases, 2,272 controls).

388 We externally validated the PRSs in two independent studies, including the NJLCC-
389 WGS[29] and the China Kadoorie Biobank (CKB). The NJLCC-WGS is a subproject of NJLCC,

390 aiming to identify rare variants of lung cancer by whole-genome sequencing. In the NJLCC-
391 WGS, 1,643 never-smoking LUAD cases (473 male; 1,170 female) and 2,235 never-smoking
392 controls (795 male; 1,440 female) were sequenced by the Illumina HiSeq or NovaSeq platforms.
393 The alignment, genotyping and quality control are previously described in detail[29]. Briefly, the
394 sequencing data from each individual were jointly genotyped for high confidence alleles using
395 GATK v3.8 according to the best practice. SNPs with inbreeding coefficients < -0.3, proportions
396 of samples with high genotype-quality (GQ) (≥ 20) ≤ 0.8, proportions of samples with enough
397 sequencing depth (≥ 10) ≤ 0.8, call rates ≤ 0.95 and Hardy-Weinberg Equilibrium p-value
398 1.0×10^{-4} were excluded.
399 The CKB is a population-based prospective cohort of 512,639 adults recruited between
400 2004 and 2008 from 10 geographical regions in China[32,33]. Due to limited histology information,
401 analyses in CKB were conducted using all never-smoking lung cancer cases (71 males; 458
402 females) rather than LUAD-specific cases. Genotyping was performed on these cases and 47,601
403 never-smoking controls (4,662 male; 42,939 female) using custom Affymetrix (now
404 ThermoFisher) Axiom® genotyping array with improved genome-wide coverage of common
405 and low-frequency variation in Chinese populations. The sample call rate threshold was 95%,
406 and the variant call rate threshold was 98%. Genotypes were imputed using the TOPMed
407 reference panel. We used the imputed genotypes with imputation INFO > 0.3.
408 Informed consent was obtained from all participants, and the recruitment and data
409 collection procedures were approved by the local ethics review committees of all participating
410 research institutes.
411
412 *Construction of PRS*

413

414 Single-ancestry PRS methods

415 **PRS-25**: We constructed a simple PRS using the 25 independent SNPs (PRS-25) that
416 reached genome-wide significance (i.e., $P<5x10^{-8}$) in the largest GWAS for LUAD in East Asian
417 population to date (**Data S1**)[8]. The PRS-25 was calculated by summing the risk alleles weighed
418 by their effect sizes (i.e., per allele log-odds ratio) obtained from a meta-analysis including only
419 never-smokers. We note of the 25 SNPs previously identified, rs4268071 (MAF = 0.04) was not
420 available in FLCCA, which we replaced with a SNP in high LD (r^2=0.84), rs4266592.

421

422 **PRS-CT:** We generated a PRS using the CT method[34] using PLINK 1.90[35]. This
423 involved the clumping of SNPs based on linkage disequilibrium (LD) with a threshold of r^2=0.1
424 within a 500kb window **Data S1**). The LD were estimated from half of the FLCCA dataset used
425 for tuning dataset. We created nine SNP subsets by applying incremental P-value thresholds (i.e.,
426 $P<5x10^{-8}$, $P<5x10^{-7}$,…, $P<5x10^{-1}$, 1), and calculated their respective PRSs using PLINK2[35]
427 command "--score cols=+scoresums,-scoreavgs no-mean-imputation". Using the tuning dataset,
428 the AUC was calculated for each threshold to identify the most predictive P-value cutoff.

429

430 **LDpred2-PRS**: We applied the LDpred2 method, implemented in the R package
431 bigsnpr[36], using a Bayesian framework to estimate SNP effect sizes through a shrinkage
432 estimator (**Data S1**). This method leverages GWAS summary statistics, incorporating a prior for
433 effect sizes, while also accounting for LD across SNPs. Our analysis was constrained to
434 HapMap3 variants, and we calibrated the model using a range of hyperparameters: the
435 proportion of causal SNPs was set across a 21-point logarithmic sequence from 10^{-5} to 1 , and

436 the per-SNP heritability was set as a fraction (0.3, 0.7, 1 or 1.4) of the total heritability estimated
437 by LD score regression[37] divided by the number of causal SNPs. Lastly, we used the "sparse"
438 option to set weaker effects to zero. The optimal tuning parameters were selected based on the
439 highest AUC achieved on the tuning dataset.
440
441 **PRS-EUR 128**: We evaluated the performance of a European-derived PRS (PRS-EUR
442 128)[38], which consists of 128 variants (**Data S1**) obtained from a GWAS of predominately active
443 smokers, on East Asian populations to assess its cross-population applicability. We note that the
444 largest GWAS conducted in European never-smokers to date have only identified a single region
445 (*TERT/CLPTM1L*) in which SNPs reached genome-wide significance[39], limiting the scope of a
446 potential targeted PRS.
447
448 Multi-ancestry PRSs methods:
449 **LDpred2 PRS + PRS-EUR 128**: For the multi-ancestry PRSs, we applied the weighted-
450 PRS approach by linearly combining the most predictive single-ancestry PRS from the East
451 Asian population (Ldpred2 PRS) with the European population-specific PRS (PRS-EUR 128)
452 (**Data S1**). The weights of this combined PRS were calculated by applying a logistic regression
453 on the tuning dataset in R version 4.2.0.
454
455 **PRS-CSx**: We applied PRS-CSx, a multi-ancestry polygenic prediction method that uses
456 a Bayesian framework with a continuous shrinkage prior to estimate SNP effect sizes from
457 GWAS summary statistics across different populations[26]. LD reference panels for European and
458 East Asian, provided by the PRS-CSx software, were constructed from the 1000 Genomes

459 Project (1KG) samples [40]. We used the default setting with the gamma-gamma prior
460 hyperparameters (a and b) at 1 and 0.5, respectively. The shrinkage parameter ϕ was assessed at
461 10^{-6}, 10^{-4}, 10^{-2}, and 1 within the tuning dataset to select the value with highest AUC. With the
462 optimal ϕ , we calculated the PRS using posterior effect-sizes from both East Asian and
463 European populations. Weights were then estimated to linearly combine the two PRSs on the
464 tuning dataset (**Data S1**). The final performance was then evaluated on an East Asian
465 independent validation dataset.
466
467 **CT-SLEB:** We used the recently developed CT-SLEB (clumping and thresholding, super
468 learning, and empirical Bayes) method to derive a PRS using East Asian and European datasets
469 [24]. This method extends the standard CT method by designing a two-dimensional approach to
470 select SNPs for East Asian PRS construction. It incorporates an empirical Bayesian (EB)
471 framework to model genetic correlations between East Asian and EUR. Following this, a super
472 learning (SL) model is then applied to integrate multiple PRSs, each generated using distinct p-
473 value thresholds and clumping parameters[37]. Our implementation of CT-SLEB followed the
474 default setting with p-value thresholds $p_t = 5\text{x}10^{-8}, 5\text{x}10^{-7}, \ldots, 5\text{x}10^{-1}$, or 1, and genetic
475 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
476 process for East Asian was based on the LD reference derived from the FLCCA tuning dataset,
477 and the European LD reference was based on the 1KG European samples.
478
479 *Relative and absolute risk calculation for PRSs*
480

481 To evaluate each PRS's performance, we standardized the scores to a unit standard
482 deviation, based on the distribution among the control group in the East Asian validation dataset.
483 We then quantified the association between the standardized PRSs and LUAD risk using the OR
484 and 95% CI, via logistic regression, adjusting for age and the first 10 principal components
485 (PCs). We evaluated the predictive performance of the different PRS models through the
486 adjusted AUC values, accounting for age and top 10 PCs, using the R package RISCA[41]. We
487 further estimated the ORs of each PRS for risk of LUAD based on PRS percentiles, setting the
488 middle quantile (40-60%) as the reference category. Lastly, we used the likelihood ratio test to
489 evaluate the interaction between the PRS and age (<40, 40-49, 50-59, 60-69, ≥70). Statistical
490 tests performed were two-sided and an alpha<0.05 was considered noteworthy.
491 For absolute risk estimates, we used the iCARE software[42] to calculate the cumulative
492 lifetime (age 30-80) and 10-year absolute risks of LUAD among never-smoking female controls
493 in FLCCA (N=4,544). Absolute risks were derived by applying the Cox proportional hazard
494 model with the top performing PRS (CT-SLEB) and age-specific lung cancer incidence and
495 mortality rates in Taiwan[43]. We chose to use incidence rates from Taiwan as approximately a
496 quarter of FLCCA samples are from Taiwan, and most individuals in Taiwan are Han Chinese,
497 similar to those in Mainland China. Notably, the lung cancer rates available for Taiwan are
498 specifically for never-smokers, which makes for more appropriate estimate of absolute risks that
499 may otherwise be inflated if ever smokers are included in the reported rates. We estimated the
500 absolute risk of LUAD overall, as well as stratified by first-degree family history of lung cancer.
501
502 *Projections of CT and LDpred2 PRS performance by sample size*
503

504 We used GENESIS package [44] to estimate the sample sizes needed for PRS to reach
505 various AUC levels in the East Asian population. This method estimates the expected number of
506 SNP discoveries and their explained heritability in future studies. Using the GWAS summary
507 statistics from our East Asian training dataset, and the provided LD scores for East Asian
508 populations from the 1KG dataset, we projected the AUC for CT PRS across various case-
509 control ratios, from 1:1 to 1:10, and case numbers ranging from 5,000 to 200,000. Given
510 GENESIS's specific design for CT PRS, we extended its application to LDpred2 PRS
511 projections. This involved modeling the relationship between effective sample sizes with the
512 phenotypic variance ratio between LDpred2 and CT PRS (**Figure S1)**, elaborated in
513 **Supplementary Methods**.
514
515 **Results**
516
517 *Development and validation of the PRS*
518
519 We applied several cutting-edge single-and multi-ancestry PRS methods (**Methods,**
520 **Figure S2**), evaluating their performance in terms of relative risk and AUC within the FLCCA
521 dataset of East Asian female never-smokers (**Methods**, **Table 1, Figure S2-S3)**. Among single-
522 ancestry PRS methods, the LDpred2 PRS, incorporating 942,591 SNPs, outperformed other
523 methods, with an odds ratio (OR) per unit standard deviation (SD) of 1.62 (95% confidence
524 interval (CI): 1.52, 1.73) alongside an adjusted AUC of 0.629 (95% CI: 0.618, 0.641). In
525 contrast, the PRS designed exclusively for European populations underperformed in our East

526 Asian population, with an estimated adjusted AUC of 0.489 (95% CI: 0.477, 0.501), possibly
527 attributed to the inclusion of smokers in the development of the existing European PRS.
528 Among multi-ancestry PRS methods, the weighted-PRS of East Asian and European
529 yielded an OR per unit SD of 1.62 (95% CI: 1.52, 1.72), with PRS-CSx and CT-SLEB showing
530 even stronger association with ORs of 1.67 (95% CI: 1.57, 1.78) and 1.71 (95% CI: 1.61, 1.82),
531 respectively. Among all the PRS methods, CT-SLEB had the highest AUC of 0.640 (95% CI:
532 0.629, 0.653). Setting individuals in the middle PRS quantile (40th to 60th percentile) as the
533 reference category, those in the highest 5% of risk for the top performing PRS, CT-SLEB, had
534 4.17 (95% CI: 3.20, 5.47)-fold risk, whereas those in the lowest 5% had 0.33 (95% CI: 0.21,
535 0.51)-fold risk of developing LUAD (**Figure 2**).
536 We further evaluated the PRSs in independent external datasets of male and female
537 never-smokers from NJLCC and CKB**.** Similar to our findings in FLCCA, we observed the
538 LDpred2 PRS had moderate performance among the single-ancestry PRS methods with an OR of
539 1.77 (95% CI: 1.61, 1.95) and an adjusted AUC of 0.645 (95% CI: 0.615, 0.672) in NJLCC
540 female never-smokers (**Table 2**). Among the multi-ancestry PRS methods, CT-SLEB yielded an
541 OR of 1.85 (95% CI: 1.68, 2.04) and an AUC of 0.649 (95% CI: 0.623, 0.678). Notably, among
542 male never-smokers, LDpred2 PRS yielded the highest relative risk estimate (OR=1.91, 95% CI:
543 1.66, 2.21) and AUC value (AUC=0.669, 95% CI: 0.633, 0.705) among both the single- and
544 multi-ancestry PRSs. However, the estimates for the different PRS approaches were not
545 statistically different. PRS performance in CKB was lower than in FLCCA and NJLCC, likely
546 due to small sample size and inablility to restrict analyses to LUAD cases only. Specifically,
547 across both the single- and multi-ancestry PRSs, PRS-25 showing the highest performance in
548 male (OR=1.29, 95% CI: 1.02, 1.63; AUC=0.611, 95% CI: 0.547, 0.673) and female (OR=1.25,

549 95% CI: 1.14, 1.37); AUC=0.572, 95% CI: 0.542, 0.606) never-smokers (**Table S1)**. However,
550 the validation was conducted among lung cancer cases with all histological subtypes, which may
551 have limited the performance of the PRSs developed for LUAD.
552
553 *Interaction between CT-SLEB PRS and age*
554
555 We observed a statistically significant multiplicative interaction between the CT-SLEB PRS and
556 age at diagnosis for LUAD risk (p-interaction = 0.002, **Figure 3**, **Table S2**), with risk estimates
557 increasing across middle-age categories (40-69) and peaking between ages 60-69 (OR = 1.93,
558 95% CI: 1.79-2.08). Individuals under 40 also exhibited an elevated PRS-associated risk (OR =
559 1.70, 95% CI: 1.42, 2.05). However, while an external validation in NJLCC was suggestive of an
560 interaction (p-interaction = 0.07), the risk estimates followed a steady increasing trend with age,
561 reaching the highest OR in individuals 70 and older (OR = 2.04, 95% CI: 1.64-2.58). When
562 stratified by sex, NJLCC results were further attenuated. External validation in CKB also did not
563 confirm an interaction between the PRS and age at diagnosis (p-interaction=0.27). While
564 findings in FLCCA suggest a potential age-related variation in PRS effects, the lack of statistical
565 significance in NJLCC and CKB indicates that this relationship may not be robust. Additional
566 studies with larger sample sizes are needed to determine whether PRS effects on LUAD risk
567 truly differ by age.
568
569 *Absolute risk of developing lung adenocarcinoma by CT-SLEB PRS*
570

571 Using the iCARE package[45], we estimated the absolute risk of LUAD for never-smoking
572 East Asian women, utilizing the CT-SLEB PRS which demonstrated the highest AUC (**Table 1**).
573 The cumulative lifetime risk of LUAD, assessed between age 30-80, varied markedly across PRS
574 percentiles, ranging from a minimal mean of 0.78% in the 5th percentile to a substantial mean of
575 6.69% in the 95th percentile (**Figure 4)**. Additionally, the 10-year absolute risk for LUAD in
576 never-smoking 50-year-old women, varied from a mean of 0.13% in the 5th percentile to mean
577 of 1.11% in the 95th percentile of the PRS distribution. While the average 10-year absolute risk
578 for this age group is 0.42%, never-smoking women in the >95 and 90-95th percentiles PRS
579 reached this risk earlier, at age of 41 and 42, respectively.
580 To evaluate the combined effect of PRS and family history, we further modeled the
581 absolute risk for LUAD in individuals with (n=86) and without (n=1,414) family history of lung
582 cancer in a first-degree relative (**Figure S4, Figure S5**). We observed that those with family
583 history of lung cancer have higher absolute lifetime risk of LUAD than those without family
584 history. The lifetime absolute risk between age 30-80 for women at the bottom 5th and top 95th
585 percentile of the PRS with family history of lung cancer were 2.77% and 9.39%, respectively
586 (**Figure S4**). Additionally, the 10-year absolute risk for women in the 95th PRS percentile with a
587 family history reached the population average 10-year absolute risk at age 50 (0.42%) much
588 earlier, at age 38. Formal tests for interaction between PRS and family history were not
589 statistically significant (p=0.45), suggesting that family history and PRS independently
590 contribute to LUAD risk stratification.
591
592 *Projections of polygenic risk score performance by sample size*
593

594 Using the GENESIS model, we projected the expected AUC of PRS-CT and LDpred2
595 PRS under varying GWAS sample sizes and case-control ratios for LUAD in never smoking East
596 Asian women (**Figure 5**). Our analysis, using the East Asian training dataset, estimated
597 approximately 1,772 (s.e.=1,641) susceptibility variants that are independently associated with
598 LUAD[27]. This high number of variants underscores the extensive polygenic nature of LUAD,
599 implying relatively small effect sizes for individual SNPs. Based on our current data, the
600 expected AUC for the LDpred2 PRS stands at 0.631 (95% CI: 0.618, 0.641), aligning closely
601 with the actual performance of the LDpred2 PRS in the validation dataset. The upper boundary
602 of our predictions, based solely on PRS, suggests an AUC potential of up to 0.731 (95% CI:
603 0.640,0.786). As sample sizes increase in our projections, the AUC is estimated to rise to 0.673
604 (95% CI: 0.593–0.725) with 55,000 East Asian LUAD cases and a 1:1 case-control ratio. With
605 the same number of cases but a 1:10 case-control ratio, the AUC is projected to further increase
606 to 0.698 (95% CI: 0.621–0.727), nearing the theoretical upper limit of predictive accuracy.
607
608 *Explaining genetic variance through PRS across different sample sizes*
609
610 We also evaluated the genetic variance explained by the PRS under different sample sizes
611 given the projected results (**Supplementary Methods, Table 3**). The current LDpred2 PRS
612 explained 26.6% of the genome-wide chip heritability, contributing to approximately 16.5% of
613 the 1.84-fold familial relative risk for lung cancer among East Asian never-smokers.
614 Concurrently, genome-wide chip variants heritability explained 61.9% of the 1.84-fold familial
615 relative risk associated with the disease. With an expanded sample size of 35,000 cases and
616 350,000 controls, the constructed PRS is projected to account for 57.9% of the genome-wide
617 chip variants heritability, and 35.8% of the 1.84-fold familial relative risk for the disease.

618

619 **Discussion**

620

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

630 To date, GWAS and subsequent PRS models have largely centered on European
631 populations [46,47], rendering them less precise when applied to non- European populations and
632 risking the exacerbation of health disparities [48,49]. Consistent with prior research [15], our study
633 observed that a lung cancer PRS constructed using European data, which included both smokers
634 and never-smokers, significantly underperformed when applied to East Asian never-smoking
635 individuals (AUC = 0.489, 95% CI: (0.477, 0.501)), further highlighting the heterogeneity of the
636 disease across populations and the need to expand risk assessment efforts to non- European
637 populations. Notably, by utilizing novel methods to integrate GWAS data from both European
638 and East Asian populations, we demonstrated that a multi-ancestry PRS enhances the precision
639 of risk stratification for LUAD among East Asian never-smokers.

640 Large-scale efforts to integrate PRS into clinical practice have focused on conditions with
641 well-established PRS prediction performance[16–18], such as Type 2 Diabetes, breast cancer, and
642 cardiovascular diseases. Lung cancer has been notably absent from these efforts, primarily due to
643 the focus on smoking as a risk factor. This has limited the exploration and application of PRS in
644 lung cancer risk stratification, particularly among never-smokers. Recent efforts, such as the
645 TALENT study in Taiwan[50], have evaluated low-dose computed tomography (LDCT) screening
646 for never-smokers with traditional risk factors like family history, passive smoking exposure, and
647 chronic lung diseases. However, a PRS has not yet been integrated into investigated screening
648 protocols. Further, a recent study in Taiwan incorporating genetic susceptibility in a risk model
649 with lifestyle and environmental risk factors observed an improvement in the AUC from 0.697 to
650 0.714[43]. Notably, the study found that even by incorporating only 11 susceptibility loci, the
651 model's discriminative power had improved. Our study demonstrates the potential of a genome-
652 wide PRS to complement these efforts, enhancing risk stratification and identifying high-risk
653 individuals at younger ages, who may benefit from earlier interventions.
654 We address this gap by presenting a PRS that is associated with a 3.92-fold increase in
655 lung cancer risk for individuals in the top 20% risk quantile, a marked improvement over the
656 2.09-fold increase observed in a prior study by Wei et al. [51] for a cohort of Chinese never-
657 smoking women. Further, we evaluated the interaction between the PRS and age in FLCCA,
658 NJLCC and CKB. The relationship between the PRS and age at diagnosis showed inconsistent
659 patterns across datasets. In FLCCA, PRS effects varied nonlinearly, peaking at ages 60-69 and
660 remaining elevated in younger individuals, suggesting a potential age-related interaction.
661 However, NJLCC showed a gradual increase in PRS-associated risk across age groups, and
662 estimates in CKB did not follow a specific patter. In both studies, the interaction was not

663 statistically significant. This discrepancy may reflect differences in sample size, population
664 structure, or environmental exposures, which could influence genetic risk expression across age
665 groups. Future studies should examine PRS-age interactions across diverse East Asian
666 populations, considering additional risk factors such as environmental and hormonal influences.

667 We estimated the lifetime and 10-year absolute risk of LUAD using the PRS, which can
668 be used in risk stratification efforts and to identify those who are at high risk for the disease.
669 Specifically, we found that the 10-year absolute risk for LUAD in a never-smoking 50-year-old
670 woman, a critical age for initiating recommended annual lung cancer screening as per the United
671 States Preventive Services Taskforce (USPSTF)[52], varied from 0.13% in the 5th percentile to
672 1.11% in the 95th percentile of the PRS distribution. Notably, women in the highest risk
673 categories (>95th and 90-95th percentiles) reached the average 10-year absolute risk (0.42%)
674 much earlier, at ages 41 and 42, respectively. This finding suggests that the PRS has the potential
675 to identify younger women who may benefit from earlier risk-based interventions. While limited
676 in sample size, we also found that the lifetime absolute risk of LUAD was even higher among
677 women in the top percentile of the PRS and with first-degree family history of lung cancer,
678 suggesting the two risk factors may act independently for their effect on LUAD risk. Although
679 the absolute risk values are modest, the PRS's ability to stratify women into higher-risk
680 categories at younger ages suggests potential utility in guiding early screening decisions. Early
681 identification through PRS could be particularly valuable for never-smoking individuals, who are
682 traditionally excluded from lung cancer screening programs.

683 The CT-SLEB method's improved performance can be attributed to its ability to utilize
684 diverse genotyping arrays, thereby including population-specific variants. This aspect, along
685 with the inclusion of European samples in the training set, broadens the genetic diversity and

686 leverages larger sample sizes for improved model accuracy. Moreover, CT-SLEB excels in
687 modeling genetic correlations across ancestries, efficiently using both shared and unique genetic
688 markers for refined disease risk estimations across populations. Lastly, we set up our study with
689 a robust three-sample split design, dividing datasets into training, tuning, and validation. This
690 ensures that PRS efficacy is validated independently, effectively reducing the overfitting risk.

691 Current PRS studies predominantly focus on evaluating relative and absolute
692 risks[8,12,15,51,53], yet our research introduces a crucial forecast of sample sizes required for
693 differing levels of PRS accuracy. Our projection analyses indicated that achieving an AUC of
694 0.70 would approach the maximized prediction potential of the PRS, given the estimated genetic
695 variance explained by GWAS chip variants. To reach this level of accuracy, a future study would
696 need to include 55,000 cases with a 1:10 case-control ratio. Notably, accessing large biobanks
697 with publicly available controls could reduce the number of required cases[54].–These projections
698 offer a strategic framework for planning and designing future genetic studies on lung cancer,
699 establishing clear benchmarks for PRS model performance, which is essential for the scientific
700 community in advancing genetic risk prediction.

701 Our study has several limitations. First, it focuses solely on the genetic susceptibility of
702 LUAD without incorporating additional questionnaire-based risk factors, especially familly
703 history. However, leveraging the largest GWAS dataset available, our primary goal was to
704 develop the most effective PRS model, which can be integrated with other risk indicators in
705 future studies. Second, our projections are based on single-ancestry PRS models and do not fully
706 account for the complexities of multi-ancestry approaches, such as varying sample sizes and
707 genetic correlations across populations. Third, our validation cohort, derived from the FLCCA
708 and covering various East Asian regions, relied on Taiwanese incidence data for absolute risk

709 estimations. Lastly, while genome-wide PRS models like LDpred2 and CT-SLEB demonstrated
710 strong performance, their reliance on nearly one million SNPs poses challenges related to
711 overfitting and clinical feasibility. Future work should aim to optimize PRS models by
712 prioritizing key loci while maintaining predictive power.

713 In summary, our study evaluates various PRS models to capture the genetic
714 predisposition to LUAD among never-smoking East Asian individuals. It extends beyond risk
715 prediction by estimating both 10-year and cumulative lifetime absolute risks, and by projecting
716 the sample sizes required for future GWAS to refine the predictive power of PRSs in future
717 GWAS. Additionally, we quantify the phenotypic variance captured by PRSs across different
718 sample sizes. Despite advancements in using PRS for smoking cessation trials[55], its
719 implementation for never-smokers in clinical settings has been limited, presenting a significant
720 area for research. Future studies are crucial to further improve these PRS models, aiming to
721 enhance genetic risk predictions while integrating a wider array of risk factors. Such efforts will
722 develop more accurate and comprehensive risk models for LUAD in never-smoking individuals
723 across different populations.
724
725
726
727

728 **Data availability statement**

729 All data supporting the findings described in this paper are available in the paper and in the
730 Supplementary Information and from the corresponding author or as otherwise indicated upon
731 request. The individual genotype data for the FLCCA data are in dbGaP phs000716.v1.p1
732 (Genome-Wide Association Study of Lung Cancer Susceptibility in Never-Smoking Women in
733 Asia). For the NCC and ACC studies, please contact Kouya Shiraishi at kshirais@ncc.go.jp or
734 Takashi Kohno at tkkohno@ncc.go.jp for summary statistics. The GWAS data for the European
735 populations contributing to this study are available at dbGap under accession phs000877.v1.p1
736 (Transdisciplinary Research Into Cancer of the Lung
737 (TRICL), https://www.ncbi.nlm.nih.gov/projects/gap/cgi-
738 bin/study.cgi?study_id=phs000876.v2.p1), phs001273.v3.p2 (Oncoarray
739 Consortium, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-
740 bin/study.cgi?study_id=phs001273.v3.p2). To gain access to all data in dbGaP cited in this
741 paper, please apply for dbGaP Authorized Access. GWAS data for the European populations
742 contributing to this study are available under accession phs000877.v1.p1 (ILLCO),
743 phs001273.v1.p1 (OncoArray).
744
745 **Code Availability**
746
747 Methods implemented and their corresponding repositories include: SCT and LDpred2
748 at https://github.com/privefl/bigsnpr, PRS-CSx at https://github.com/getian107/PRScsx. CT-SLEB
749 at https://github.com/andrewhaoyu/CTSLEB, and GENESIS at
750 https://github.com/yandorazhang/GENESIS.
751 PLINK: https://www.cog-genomics.org/plink/1.9.
752 Most of our statistical analyses were performed using the following R packages: ggplot2 v.3.3.3,
753 dplyr v.1.0.4, data.table v.1.13.6, iCARE v.1.30.0
754 https://github.com/KevinWFred/PRS_EASLC
755
756 **Funding**

757 Female Lung Cancer Consortium in Asia (NCI**):** This study was supported by a Grant-in-Aid for
758 Scientific Research on Priority Areas from the Ministry of Education, Science, Sports, Culture
759 and Technology of Japan, a Grant-in- Aid for the Third Term Comprehensive 10-Year Strategy
760 for Cancer Control from the Ministry Health, Labor and Welfare of Japan, by Health and Labor
761 Sciences Research Grants for Research on Applying Health Technology from the Ministry of
762 Health, Labor and Welfare of Japan, by the National Cancer Center Research and Development
763 Fund, the National Research Foundation of Korea (NRF) grant funded by the Korea government
764 (MEST) (grant No. 2011-0016106), a grant of the National Project for Personalized Genomic
765 Medicine, Ministry for Health & Welfare, Republic of Korea (A111218-11-GM04), the Program
766 for Changjiang Scholars and Innovative Research Team in University in China (IRT_14R40 to
767 K.C.), the National Science & Technology Pillar Program (2011BAI09B00), MOE 111 Project
768 (B13016), the National Natural Science Foundation of China (No. 30772531, and 81272618),
769 Guangdong Provincial Key Laboratory of Lung Cancer Translational Medicine (No.
770 2012A061400006), Special Fund for Research in the Public Interest from the National Health
771 and Family Planning Commission of PRC (No. 201402031), and the Ministry of Science and
772 Technology, Taiwan (MOST 103-2325-B-400-023 & 104-2325-B-400-012). The Japan Lung
773 Cancer Study (JLCS) was supported in part by the Practical Research for Innovative Cancer
774 Control from Japan Agency for Medical Research and Development (15ck0106096h0002) and
775 the Management Expenses Grants from the Government to the National Cancer Center (26-A-1)
776 for Biobank. BioBank Japan was supported by the Ministry of Education, Culture, Sports,
777 Sciences and Technology of the Japanese government. The Japan Public Health Center-based
778 prospective Study (the JPHC Study) was supported by the National Cancer Center Research and
779 Development Fund (23-A- 31[toku], 26-A-2, 29-A-4, 2020-J-4, and 2023-J-4) (since 2011) and a
780 Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan
781 (from 1989 to 2010). The Taiwan GELAC Study (Genetic Epidemiological Study for Lung
782 AdenoCarcinoma) was sup- ported by grants from the National Research Program on Genomic
783 Medicine in Taiwan (DOH99-TD-G-111-028), the National Research Program for
784 Biopharmaceuticals in Taiwan (MOHW 103-TDUPB-211-144003, MOST 103-2325-B-400-023)
785 and the Bioinformatics Core Facility for Translational Medicine and Biotechnology
786 Development (MOST 104-2319-B-400-002). This work was also supported by the Jinan Science
787 Research Project Foundation (201102051), the National Key Scientific and Technological

788 Project (2011ZX09307-001-04), the National Natural Science Foundation of China
789 (No.81272293), the State Key Program of National Natural Science of China (81230067), the
790 National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP)
791 (No. NRF- 2014R1A2A2A05003665), Sookmyung Women's University Research Grants, Korea
792 (1-1603-2048), Agency for Science, Technology and Research (A*STAR), Singapore and the
793 US National Institute of Health Grant (1U19CA148127-01). The overall GWAS project was
794 supported by the intramural program of the US National Institutes of Health/National Cancer
795 Institute. The following is a list of grants by study center: SKLCS (Y.T.K.)—National Research
796 Foundation of Korea (NRF) grant funded by the Korea government (MEST) (2011-0016106).
797 (J.C.) – This work was supported by a grant from the National R&D Program for Cancer
798 Control, Ministry of Health &Welfare, Republic of Korea (grant no. 0720550-2). (J.S.S) – grant
799 number is A010250. WLCS (T.W.)—National Key Basic Research and Development Program
800 (2011CB503800). SLCS (B.Z.)—National Nature Science Foundation of China (81102194).
801 Liaoning Provincial Department of Education (LS2010168). China Medical Board (00726). GDS
802 (Y.L.W.)—Foundation of Guangdong Science and Technology Department (2006B60101010,
803 2007A032000002, 2011A030400010). Guangzhou Science and Information Technology Bureau
804 (2011Y2-00014). Chinese Lung Cancer Research Foundation, National Natural Science
805 Foundation of China (81101549). Natural Science Foundation of Guangdong Province
806 (S2011010000792). TLCS (K.C., B.Q)—Program for Changjiang Scholars and Innovative
807 Research Team in University (PCSIRT), China (IRT1076). Tianjin Cancer Institute and Hospital.
808 National Foundation for Cancer Research (US). FLCS (J.C.W., D.R., L.J.)—Ministry of Health
809 (201002007). Ministry of Science and Technology (2011BAI09B00). National S&T Major
810 Special Project (2011ZX09102-010-01). China National High-Tech Research and Development
811 Program (2012AA02A517, 2012AA02A518). National Science Foundation of China
812 (30890034). National Basic Research Program (2012CB944600). Scientific and Technological
813 Support Plans from Jiangsu Province (BE2010715). NLCS (H.S.)—China National High-Tech
814 Research and Development Program Grant (2009AA022705). Priority Academic Program
815 Development of Jiangsu Higher Education Institution. National Key Basic Research Program
816 Grant (2011CB503805). GEL-S (A.S.)—National Medical Research Council Singapore grant
817 (NMRC/0897/2004, NMRC/1075/2006). (J.Liu)—Agency for Science, Technology and
818 Research (A*STAR) of Singapore. GELAC (C.A.H.)—National Research Program on Genomic

819 Medicine in Taiwan (DOH98-TDG-111-015). National Research Program for
820 Biopharmaceuticals in Taiwan (DOH 100- TD-PB-111-TM013). National Science Council,
821 Taiwan (NSC 100- 2319-B-400-001). YLCS (Q.L.)—Supported by the intramural pro- gram of
822 U.S. National Institutes of Health, National Cancer Institute. SWHS (W.Z., W-H.C., N.R.)—The
823 work was supported by a grant from the National Institutes of Health (R37 CA70867, UM1
824 CA182910) and the National Cancer Institute intramural research program, including NCI
825 Intramural Research Program contract (N02 CP1101066). JLCS (K.M., T.K.)—Grants-in-Aid
826 from the Ministry of Health, Labor, and Welfare for Research on Applying Health Technology
827 and for the 3rd-term Comprehensive 10-year Strategy for Cancer Control; by the National
828 Cancer Center Research and Development Fund; by Grant-in-Aid for Scientific Research on
829 Priority Areas and on Innovative Area from the Ministry of Education, Science, Sports, Culture
830 and — Technology of Japan. (W.P.)—NCI R01-CA121210. HKS (J.W.)— General Research
831 Fund of Research Grant Council, Hong Kong (781511M). The Environment and Genetics in
832 Lung Cancer Etiology (EAGLE), Prostate, Lung, Colon, Ovary Screening Trial (PLCO), and
833 Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) studies were supported by the
834 Intramural Research Program of the National Institutes of Health, National Cancer Institute
835 (NCI), Division of Cancer Epidemiology and Genetics. ATBC was also supported by U.S. Public
836 Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-RC-37004) from the NCI.
837 PLCO was also supported by individual contracts from the NCI to the University of Colorado
838 Denver (NO1-CN-25514), Georgetown University (NO1-CN-25522), the Pacific Health
839 Research Institute (NO1-CN-25515), the Henry Ford Health System (NO1-CN-25512), the
840 University of Minnesota, (NO1-CN- 25513), Washington University (NO1-CN-25516), the
841 University of Pittsburgh (NO1-CN-25511), the University of Utah (NO1-CN- 25524), the
842 Marshfield Clinic Research Foundation (NO1-CN- 25518), the University of Alabama at
843 Birmingham (NO1-CN- 75022), Westat, Inc. (NO1-CN-25476), and the University of California,
844 Los Angeles (NO1-CN-25404). The Carotene and Retinol Efficacy Trial (CARET) is funded by
845 the National Cancer Institute, National Institutes of Health through grants U01-CA063673,
846 UM1-CA167462, and U01-CA167462. The Cancer Prevention Study-II (CPS-II) Nutrition
847 Cohort was supported by the American Cancer Society. The NIH Genes, Environment and
848 Health Initiative (GEI) partly funded DNA extraction and statis- tical analyses (HG-06-033-NCI-
849 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.

Female Lung Cancer Consortium in Asia (Taiwan): 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), MOST 111-2740-B-400-002 (CAH), MOST 111-2314-B-400-020 (CAH).

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).

This research was also supported by the National Natural Science Foundation of China (82388102)

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.

This research was supported [in part] by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Conflict of Interest

The authors declare no competing interests.

Acknowledgements

We gratefully acknowledge the important historical contribution that the following individuals made to the development and success of this long-term collaboration:

Minsun Song, PhD, Department of Statistics & Research Institute of Natural Sciences, Sookmyung Women's University, Seoul, Republic of Korea;

Michiaki Kubo, PhD, Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, , Japan;

Haruhiko Nakayama, PhD, Department of Thoracic Surgery, Kanagawa Cancer Center, Yokohama, Japan;

Yuichiro Ohe, PhD,Department of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan;

Jian Su, PhD, 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;

Ying-Huang Tsai, PhD, Department of Respiratory Therapy, Chang Gung University, Taoyuan, Taiwan;

Ming-Shyan Huang, PhD, Department of Internal Medicine, E-Da Cancer Hospital, I-Shou University and Kaohsiung Medical University, Kaohsiung, Taiwan;

Kun-Chieh Chen, PhD, Department of Internal Medicine, Division of Pulmonary Medicine, Chung Shan Medical University Hospital, Taichung, Taiwan;

Biyun Qian, PhD, Department of Epidemiology and Biostatistics, Key Laboratory of Prevention and Control of Human Major Diseases of the Ministry of Education, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China;

Chen Wu, PhD, 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;

Margaret R Spitz, PhD, Department of Medicine, Section of Epidemiology and Population Science, Institute for Clinical and Translational Research, Houston, TX, USA;

Mikael Johansson, PhD, Department of Radiation Sciences, Umeå University, Umeå, Sweden;

Jae Sook Sung, PhD, Department of Internal Medicine, Division of Oncology/Hematology, College of Medicine, Korea University Anam Hospital, Seoul, Republic of Korea;

Yoo Jin Jung, PhD, Department of Thoracic and Cardiovascular Surgery, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea;

Huan Guo, PhD, 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;

Bin Zhu, PhD, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA;

Li Liu, PhD, Department of Oncology, Cancer Center, Union Hospital, Huazhong University of Science and Technology, Wuhan, China;

Ying Chen, PhD, Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore;

Jun Suk Kim, PhD, Department of Internal Medicine, Division of Medical Oncology, College of Medicine, Korea University Guro Hospital, Seoul , Republic of Korea;

Ho-Il Yoon, PhD,Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea;

Ping Xu, PhD,Department of Oncology, Wuhan Iron and Steel (Group) Corporation Staff-Worker Hospital , Wuhan, China;

Chih-Liang Wang, PhD, Department of Thoracic Medicine, Division of Pulmonary Oncology and Interventional Bronchoscopy, Chang Gung Memorial Hospital, Taoyuan, Taiwan;

Wei-Yen Lim, PhD, Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore;

Hongyan Chen, PhD, Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China;

Li Jin, PhD, Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China;

Shengchao A. Li, MS, Division of Cancer Epidemiology and Genetics, National Cancer Institute, , Rockville, MD, USA;

Gening Jiang, PhD, Shanghai Pulmonary Hospital, Shanghai, China;

Ke Fei, PhD, Shanghai Pulmonary Hospital, Shanghai, , China;

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

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-CLPTM1Ll region. *J. Thorac Oncol .* **14**, 1360–1369 (2019).
14. Lebrett, 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 multiancestry 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 2023 14:1* **14**, 1–17 (2023).
28. 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 (2012).
29. Wang, C. *et al.* Analyses of rare predisposing variants of lung cancer in 6,004 whole genomes in Chinese. *Cancer Cell* **40**, 1223-1239.e6 (2022).
30. Dai, J. *et al.* Risk loci identification and polygenic risk score in prediction of lung cancer: a large-scale prospective cohort study in Chinese. *Lancet Respir Med* **7**, 881 (2019).
31. 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).
32. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* **40**, 1652–1666 (2011).
33. Walters, R. G. *et al.* Genotyping and population characteristics of the China Kadoorie Biobank. *Cell Genomics* **3**, 100361 (2023).
34. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
35. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
36. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).
37. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
38. Hung, R. J. *et al.* Assessing Lung Cancer Absolute Risk Trajectory based on a Polygenic Risk Model. *Cancer Res* **81**, 1607 (2021).
39. Hung, R. J. *et al.* LUNG CANCER RISK IN NEVER SMOKERS OF EUROPEAN DESCENT IS ASSOCIATED WITH GENETIC VARIATION IN THE 5P15.33 TERT-CLPTM1L REGION. *J Thorac Oncol* **14**, 1360 (2019).
40. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

41. 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).
42. Pal Choudhury, P. *et al.* iCARE: An R package to build, validate and apply absolute risk models. *PLoS One* **15**, e0228198 (2020).
43. Chien, L. H. *et al.* Predicting lung cancer occurrence in never-smoking females in Asia: TNSF-SQ, a prediction model. *Cancer Epidemiology Biomarkers and Prevention* **29**, 452–459 (2020).
44. 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).
45. Choudhury, P. P. *et al.* iCARE: An R package to build, validate and apply absolute risk models. *PLoS One* **15**, (2020).
46. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
47. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, (2019).
48. Kullo, I. J. *et al.* Polygenic scores in biomedical research. *Nat. Rev. Genet.* **23**, 524–532 (2022).
49. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
50. Chang, G. C. *et al.* Low-dose CT screening among never-smokers with or without a family history of lung cancer in Taiwan: a prospective cohort study. *Lancet Respir Med* **12**, 141–152 (2024).
51. 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).
52. Krist, A. H. *et al.* Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA - Journal of the American Medical Association* **325**, 962–970 (2021).
53. 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).
54. 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).
55. Chen, T. *et al.* Genomic insights for personalised care in lung cancer and smoking cessation: motivating at-risk individuals toward evidence-based health practices. *EBioMedicine* **110**, 105441 (2024).

1045 **Table 1. Prediction performance of different methods for generating polygenic risk scores**
1046 **for lung cancer in never-smoking East Asian populations in FLCCA.**

Model type	Number of SNPs	OR (95% CI)[a]	AUC (95% CI)[b]
Single-ancestry method[c]			
PRS-25[d]	25	1.52 (1.43, 1.61)	0.618 (0.613, 0.636)
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[e]	128 (109 available)	0.97 (0.91, 1.03)	0.489 (0.477, 0.501)
Multi-ancestry method[f]			
LDpred2 PRS + PRS-EUR 128	942,681	1.62 (1.52, 1.72)	0.629 (0.617, 0.640)
PRS-CSx	1,031,786	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)

1047 Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; AUC, area
1048 under the curve; PC, principal component; CT, clumping and thresholding
1049 [a] ORs are provided per one unit standard deviation of PRS.
1050 [b] Adjusted for age (continuous) and first 10 principal components.
1051 [c] Sample size for single-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian
1052 ancestry for training, 2,219 cases and 2,272 controls of East Asian ancestry for tuning and 2,219 cases
1053 and 2,272 controls of East Asian ancestry for validation.
1054 [d] SNPs that reached genome-wide significance in Shi et al., 2023.
1055 [e] SNPs obtained from genome-wide association study conducted in individuals with European ancestry
1056 (Hung et al., 2021, Cancer Res.).
1057 [f] Sample size for multi-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian
1058 ancestry, as well as 2,058 cases and 5,575 controls of European ancestry for training, 2,219 cases and
1059 2,272 controls of East Asian ancestry for tuning and 2,219 cases and 2,272 controls of East Asian
1060 ancestry for validation.
1061

Table 2. Prediction performance of different methods for generating polygenic risk scores for lung cancer in never-smoking East Asian populations from Nanjing, China.

		Overall		Male		Female	
Model type	**Number of SNPs**	**OR (95% CI)[a]**	**AUC (95% CI)[b]**	**OR (95% CI)[a]**	**AUC (95% CI)[b]**	**OR (95% CI)[a]**	**AUC (95% CI)[b]**
Single-ancestry method[c]							
PRS-25[d]	25	1.51 (1.40, 1.63)	0.615 (0.592, 0.638)	1.50 (1.32, 1.73)	0.616 (0.576, 0.654)	1.52 (1.38, 1.67)	0.602 (0.575, 0.632)
PRS-CT	8	1.33 (1.23, 1.43)	0.574 (0.549, 0.600)	1.40 (1.23, 1.60)	0.593 (0.553, 0.637)	1.30 (1.19, 1.42)	0.561 (0.529, 0.59)
LDpred2 PRS	942,591	1.80 (1.67, 1.95)	0.657 (0.635, 0.678)	1.91 (1.66, 2.21)	0.669 (0.633, 0.705)	1.77 (1.61, 1.95)	0.645 (0.615, 0.672)
PRS-EUR 128[e]	128 (109 available)	1.04 (0.96, 1.12)	0.507 (0.481, 0.533)	1.05 (0.92, 1.20)	0.521 (0.473, 0.561)	1.03 (0.94, 1.12)	0.506 (0.473, 0.538)
Multi-ancestry method[f]							
LDpred2 PRS + PRS-EUR 128	942,681	1.79 (1.65, 1.94)	0.656 (0.632, 0.675)	1.91 (1.66, 2.21)	0.668 (0.631, 0.700)	1.75 (1.59, 1.93)	0.642 (0.617, 0.67)
PRS-CSx	1,031,786	1.81 (1.67, 1.96)	0.654 (0.630, 0.676)	1.87 (1.63, 2.16)	0.659 (0.619, 0.694)	1.81 (1.64, 1.99)	0.645 (0.613, 0.672)
CT-SLEB	2,127,229	1.78 (1.64, 1.92)	0.653 (0.631, 0.675)	1.71 (1.49, 1.96)	0.653 (0.618, 0.692)	1.85 (11.68, 2.04)	0.649 (0.623, 0.676)

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

[a] ORs are provided per one unit standard deviation of PRS.

[b] Adjusted for age (continuous) and first 10 principal components.

[c] Sample size for single-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian ancestry for training, 2,219 cases and 2,272 controls of East Asian ancestry for tuning and 1,643 never-smoking LUAD cases (473 male; 1,170 female) and 2,235 never-smoking controls (795 male; 1,440 female) of East Asian ancestry for validation.

[d] SNPs that reached genome-wide significance in Shi et al., 2023.

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

[f] Sample size for multi-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian ancestry, as well as 2,058 cases and 5,575 controls of European ancestry for training, 2,219 cases and 2,272 controls of East Asian ancestry for tuning and 1,643 never-smoking LUAD cases (473 male; 1,170 female) and 2,235 never-smoking controls (795 male; 1,440 female) of East Asian ancestry for validation.

Table 3. Genetic variance in East Asian lung cancer among never smokers explained by LDpred2 PRS

Sample size[a]	Genetic variance of PRS[b]	Proportion of all-GWAS variants genetic variance explained by PRS[c]	Proportion of Familial risk explained by PRS[d]
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%

[a] *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 Methods**).

[b] *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.

[c] *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, β_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.

[d] 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 λ_s is the familial risk when a first-order sibling has the disease, and σ^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 Legends

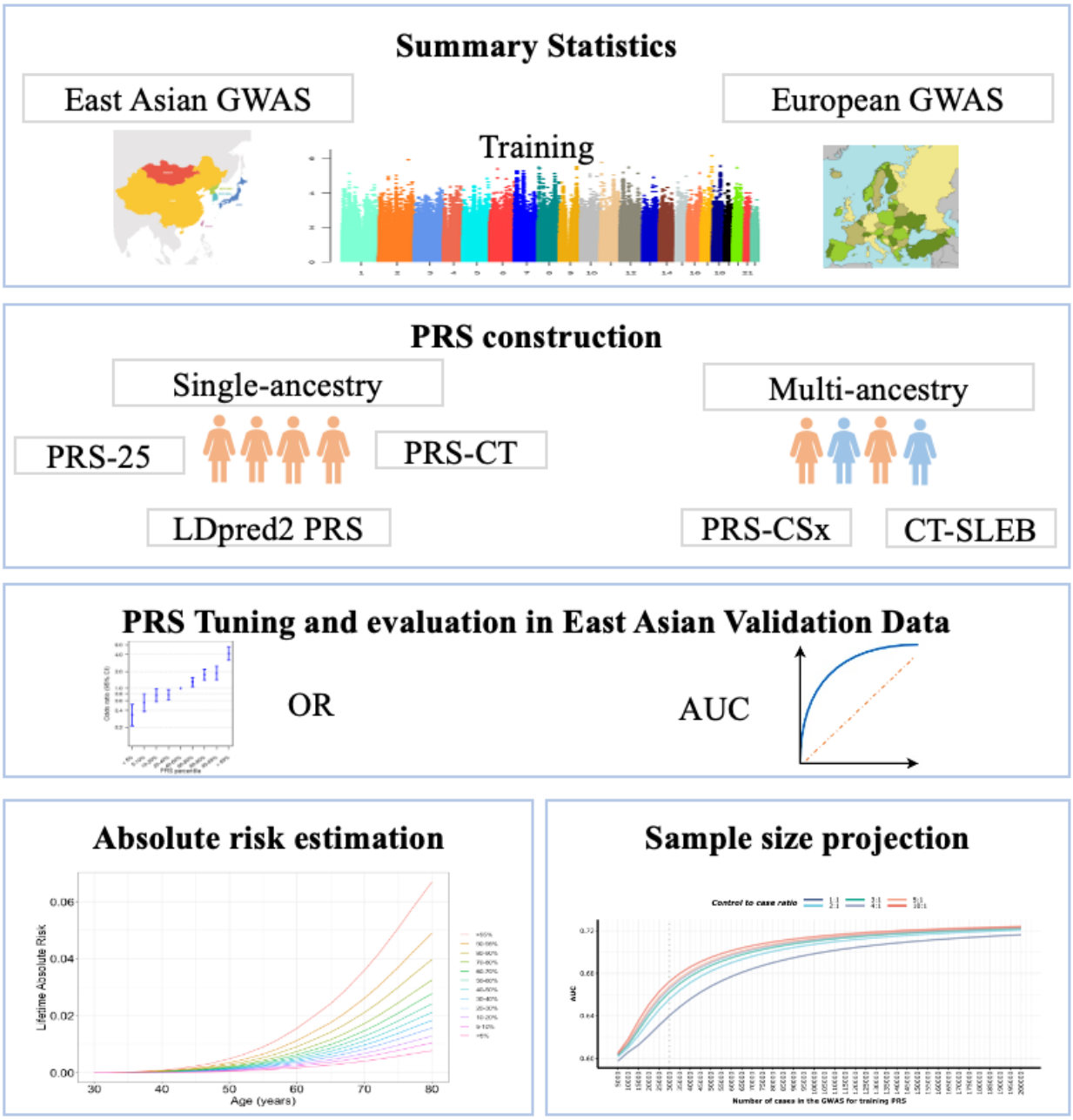


Figure 1. Overview of data structure, polygenic risk score (PRS) development, validation and application. Summary statistics from East Asian genome-wide association studies (GWAS) were used to develop single-ancestry PRS using methods such as a simple PRS constructed using 25 SNPs that have previously 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 East Asian PRS construction by incorporating over 2 million SNPs. Tuning and validation of each PRS was conducted in an independent East Asian 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.

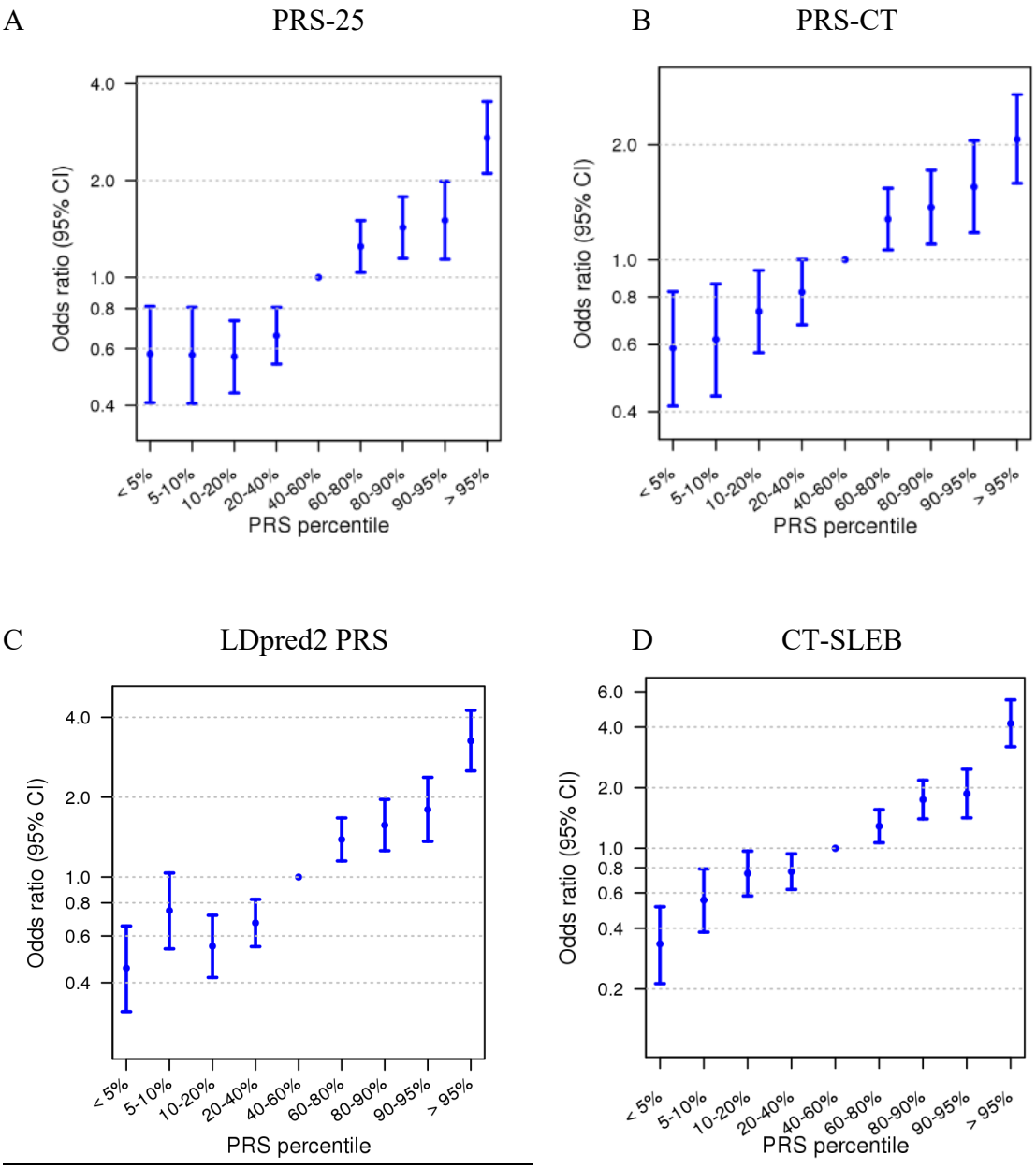
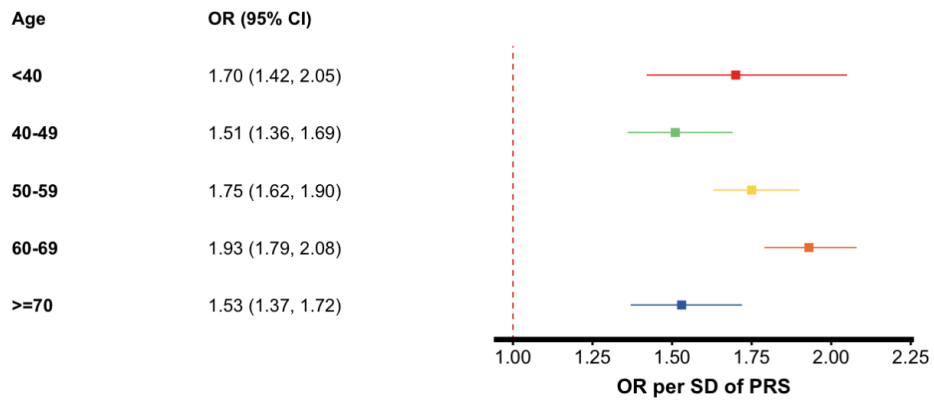


Figure 2. Relative risk estimated for quantiles of each polygenic risk score (PRS) and lung adenocarcinoma in the validation dataset of women with East Asian ancestry, treating the 40th to 60th percentile as the references. Odds ratios of PRS per standard deviation (SD) and 95% confidence intervals are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

A



B

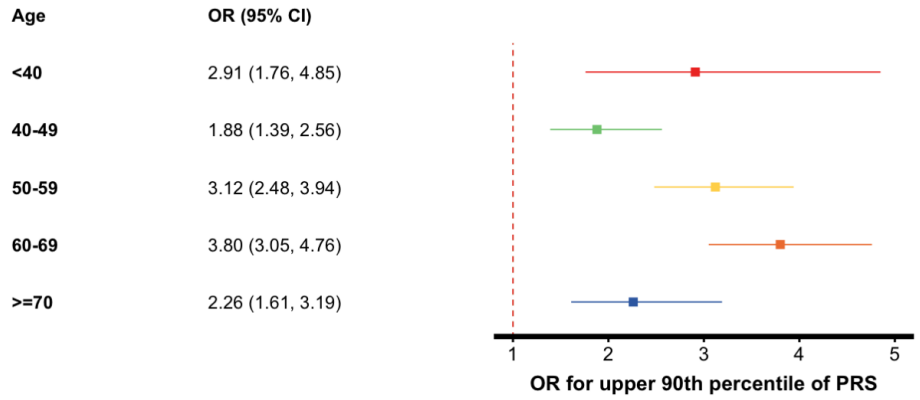
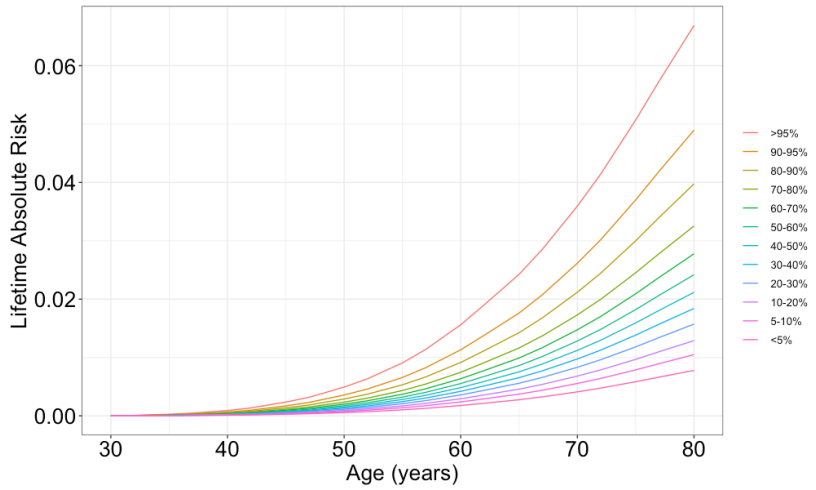


Figure 3. Association between polygenic risk score (PRS) and lung adenocarcinoma by age groups. (A) Odds ratios (ORs) per standard deviation (SD) of the PRS and 95% confidence intervals, and (B) ORs for individuals in the upper 90th percentile of the PRS.

A



B

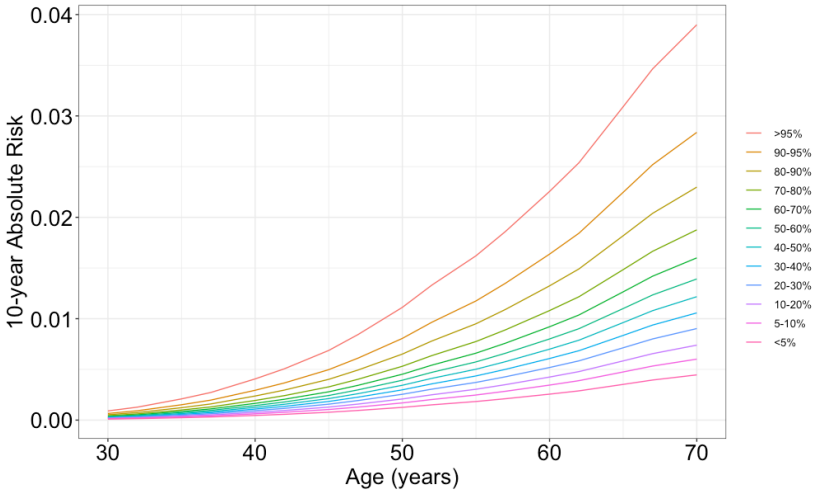
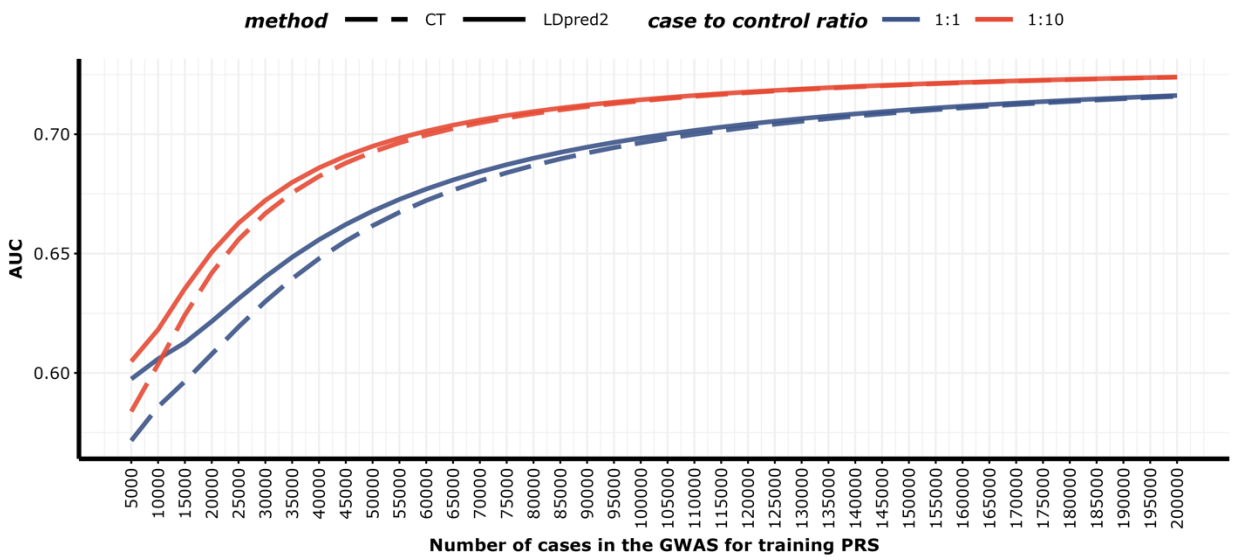


Figure 4. Lifetime cumulative and 10-year absolute risk of developing lung adenocarcinoma. (A) Lifetime (age 30-80) cumulative risk and (B) 10-year absolute risk of developing lung adenocarcinoma in never-smoking women in East Asia by percentiles of the CT-SLEB polygenic risk score (PRS). Absolute risks were calculated using the iCARE package[42], based on Taiwan's age-specific incidence and mortality data, and the PRS relative risks, as described in the Methods section.

A



B

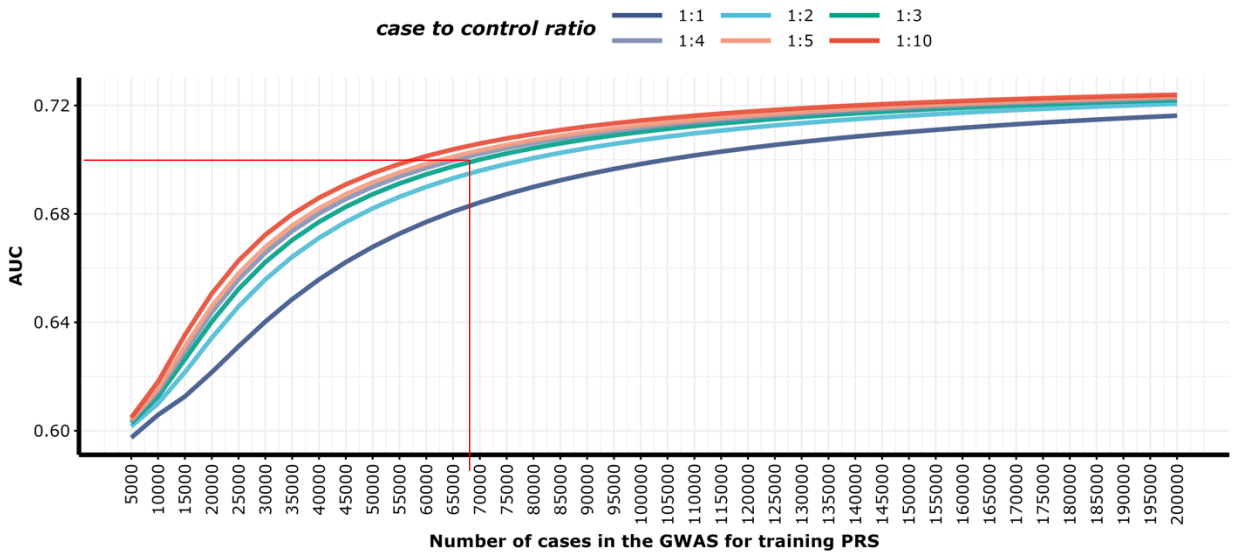


Figure 5. Projected area under the receiver operating characteristic curve (AUC) of polygenic risk scores (PRS) built using genome-wide association studies (GWAS) with varying sample sizes for lung adenocarcinoma in never-smoking East Asian women. (A) AUC values for PRS-CT and LDpred2 PRS with case-to-control ratios of 1:1 and 1:10. (B) AUC values for LDpred2 PRS across case-to-control ratios of 1:1, 1:2, 1:3, 1:4, 1:5, and 1:10

Author contributions

B.B., N.R., Q.L. and H.Z. organized and designed the study. B.B., X.W., H.Z. carried out the statistical analyses with supervision from N.R., Q.L. and H.Z. C.K. conducted the validation analyses in CKB. J.D., Y.W., T. Q. conducted the validation analyses in NJLCC. B.B., X.W., Q.L. N.R. and H.Z. wrote the first draft. B.B., X.W., C.K., J.D. Q.L. N.R. and H.Z. review and edit the draft. Q.L., M.T.L., B.A.B., W.H., 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, 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, SJ.A., XC.Z., J.S., YL.W., 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., 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., Jie.L, Z.W., S.C., J.Y., JC.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., 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., 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., 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., 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, S.W., Y.Yatabe, M.T., R.Hamamoto, A.Takahashi, Y.Momozawa, M.Kubo, Y.K., Y.D., Y.Miyagi, H.N., 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., 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., 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., 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., O.Y.G., I.P.G., X.W., P.Y. conducted epidemiology studies and contributed samples to GWAS and/or conducted initial genotyping. All authors reviewed and approved the final review of the manuscript.