1 Stratifying Lung Adenocarcinoma Risk with Multi-ancestry Polygenic Risk Scores in East

2 Asian Never-Smokers

3

4 Batel Blechter^{1*}, Xiaoyu Wang^{1*}, Jianxin Shi^{1*}, Kouya Shiraishi^{2*}, Jiyeon Choi^{1*}, Keitaro 5 Matsuo^{3*}, Tzu-Yu Chen^{4*}, Juncheng Dai^{5,6*}, Rayjean J Hung^{7*}, Kexin Chen^{8*}, Xiao-Ou Shu^{9*}, Young Tae Kim¹⁰, Parichoy Pal Choudhury¹¹, Jacob Williams¹, Maria Teresa Landi¹, Dongxin 6 Lin¹², Wei Zheng⁹, Zhihua Yin¹³, Baosen Zhou¹⁴, Jiucun Wang^{15,16}, Wei Jie Seow^{1,17,18}, Lei 7 Song¹, I-Shou Chang¹⁹, Wei Hu¹, Li-Hsin Chien^{4,20}, Qiuyin Cai⁹, Yun-Chul Hong²¹, Hee Nam 8 Kim²², Yi-Long Wu²³, Maria Pik Wong²⁴, Brian Douglas Richardson^{1,25}, Shilan Li^{1,26}, Tongwu 9 Zhang¹, Charles Breeze¹, Zhaoming Wang²⁷, Bryan A Bassig²⁸, Jin Hee Kim²⁹, Demetrius 10 Albanes¹, Jason YY Wong³⁰, Min-Ho Shin²¹, Lap Ping Chung²⁴, Yang Yang³¹, She-Juan An²³, 11 Hong Zheng⁸, Yasushi Yatabe³², Xu-Chao Zhang²³, Young-Chul Kim^{33,34}, Neil E Caporaso¹, 12 Jiang Chang³⁵, James Chung Man Ho³⁶, Michiaki Kubo³⁷, Yataro Daigo^{38,39}, Minsun Song⁴⁰, 13 Yukihide Momozawa³⁷, Yoichiro Kamatani⁴¹, Masashi Kobayashi⁴², Kenichi Okubo⁴², Takayuki 14 Honda⁴³, H Dean Hosgood⁴⁴, Hideo Kunitoh⁴⁵, Shun-ichi Watanabe⁴⁶, Yohei Miyagi⁴⁷, Haruhiko 15 Nakayama⁴⁸, Shingo Matsumoto⁴⁹, Hidehito Horinouchi⁴⁶, Masahiro Tsuboi⁵⁰, Ryuji 16 Hamamoto⁵¹, Koichi Goto⁴⁹, Yuichiro Ohe⁴⁶, Atsushi Takahashi⁴¹, Akiteru Goto⁵², Yoshihiro 17 Minamiya⁵³, Megumi Hara⁵⁴, Yuichiro Nishida⁵⁴, Kenji Takeuchi⁵⁵, Kenji Wakai⁵⁵, Koichi 18 Matsuda⁵⁶, Yoshinori Murakami⁵⁷, Kimihiro Shimizu⁵⁸, Hiroyuki Suzuki⁵⁹, Motonobu Saito⁶⁰, 19 Yoichi Ohtaki⁶¹, Kazumi Tanaka⁶¹, Tangchun Wu⁶², Fusheng Wei⁶³, Hongji Dai⁸, Mitchell J 20 Machiela¹, Jian Su²³, Yeul Hong Kim⁶⁴, In-Jae Oh^{33,34}, Victor Ho Fun Lee⁶⁵, Gee-Chen 21 Chang^{66,67,68,69}, Ying-Huang Tsai^{70,71}, Kuan-Yu Che⁷², Ming-Shyan Huang⁷³, Wu-Chou Su⁷⁴, 22 Yuh-Min Chen⁷⁵, Adeline Seow¹⁷, Jae Yong Park⁷⁶, Sun-Seog Kweon^{21,77}, Kun-Chieh Chen⁶⁷, 23 Yu-Tang Gao⁷⁸, Biyun Qian⁸, Chen Wu¹², Daru Lu^{15,16}, Jianjun Liu^{76,79,80}, Ann G Schwartz⁸¹, 24 Richard Houlston⁸², Margaret R Spitz⁸³, Ivan P Gorlov⁸³, Xifeng Wu⁸⁴, Ping Yang⁸⁵, Stephen 25 Lam⁸⁶, Adonina Tardon⁸⁷, Chu Chen⁸⁸, Stig E Bojesen^{89,90}, Mattias Johansson⁹¹, Angela 26 27 Risch^{92,93,94}, Heike Bickeböller⁹⁵, Bu-Tian Ji¹, H-Erich Wichmann^{96,97,98}, David C. Christiani⁹⁹, Gadi Rennert¹⁰⁰, Susanne Arnold¹⁰¹, Paul Brennan⁹¹, James McKay⁹¹, John K Field¹⁰², Michael 28 P.A. Davies¹⁰², Sanjay S Shete¹⁰³, Loic Le Marchand¹⁰⁴, Geoffrey Liu¹⁰⁵, Angeline Andrew¹⁰⁶, 29 Lambertus A Kiemeney¹⁰⁷, Shan Zienolddiny-Narui¹⁰⁸, Kjell Grankvist¹⁰⁹, Mikael Johansson¹¹⁰, 30 Angela Cox¹¹¹, Fiona Taylor¹¹¹, Jian-Min Yuan¹¹², Philip Lazarus¹¹³, Matthew B Schabath¹¹⁴, 31 Melinda C Aldrich¹¹⁵, Hyo-Sung Jeon¹¹⁶, Shih Sheng Jiang¹⁹, Jae Sook Sung⁶⁴, Chung-Hsing 32 Chen¹⁹, Chin-Fu Hsiao⁴, Yoo Jin Jung¹¹⁷, Huan Guo¹¹⁸, Zhibin Hu⁵, Laurie Burdett^{1,119}, Meredith 33 Yeager^{1,119}, Amy Hutchinson^{1,119}, Belynda Hicks^{1,119}, Jia Liu^{1,119}, Bin Zhu^{1,119}, Sonja I Berndt¹, 34 Wei Wu¹³, Junwen Wang^{120,121}, Yuqing Li⁷⁹, Jin Eun Choi¹¹⁶, Kyong Hwa Park⁶⁴, Sook Whan 35 Sung¹²², Li Liu¹²³, Chang Hyun Kang¹¹⁷, Wen-Chang Wang¹²⁴, Jun Xu¹²⁵, Peng Guan^{13,126}, Wen 36 Tan¹², Chong-Jen Yu¹²⁷, Gong Yang⁹, Alan Dart Loon Sihoe¹²⁸, Ying Chen¹⁷, Yi Young Choi¹¹⁶, 37 Jun Suk Kim¹²⁹, Ho-Il Yoon¹³⁰, In Kyu Park¹¹⁷, Ping Xu¹³¹, Qincheng He¹³, Chih-Liang 38 Wang¹³², Hsiao-Han Hung^{19,133}, Roel C.H. Vermeulen¹³⁴, Iona Cheng¹³⁵, Junjie Wu^{15,16}, Wei-39 Yen Lim¹⁷, Fang-Yu Tsai^{19,133}, John K.C. Chan¹³⁶, Jihua Li¹³⁷, Hongyan Chen^{15,16}, Hsien-Chih 40 Lin⁴, Li Jin^{15,16}, Jie Liu¹³⁸, Norie Sawada¹³⁹, Taiki Yamaji¹⁴⁰, Kathleen Wyatt^{1,119}, Shengchao A. 41 Li^{1,119}, Hongxia Ma^{5,6}, Meng Zhu^{5,6}, Zhehai Wang¹³⁸, Sensen Cheng¹³⁸, Xuelian Li^{13,126}, Yangwu 42 Ren^{13,126}, Ann Chao¹⁴¹, Motoki Iwasaki^{139,140}, Junjie Zhu³⁰, Gening Jiang³⁰, Ke Fei³⁰, Guoping 43 Wu⁶³. Chih-Yi Chen¹⁴², Chien-Jen Chen¹⁴³, Pan-Chyr Yang¹⁴⁴, Jinming Yu¹³⁸, Victoria L. 44 Stevens¹⁴⁵, Joseph F. Fraumeni Jr¹, Nilanjan Chatterjee^{1,146,147‡}, Olga Y Gorlova^{148,149,150‡}, 45

Christopher I Amos^{148,149,150‡}, Hongbing Shen^{5,6‡}, Chao Agnes Hsiung^{4‡}, Stephen J Chanock^{1‡},
Nathaniel Rothman^{1‡}, Takashi Kohno^{2‡}, Qing Lan^{1‡}, Haoyu Zhang^{1‡}

49 ¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, 50 USA, ²Division of Genome Biology, National Cancer Research Institute, Tokyo, Japan, 51 ³Division of Molecular Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan, 52 ⁴Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan, 53 ⁵Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, 54 China, ⁶Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China, 55 56 ⁷Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada, ⁸Department of Epidemiology and Biostatistics, National Clinical 57 Research Center for Cancer, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, 58 59 Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, 60 China, ⁹Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center and Vanderbilt-Ingram Cancer Center, Nashville, TN, USA, ¹⁰Cancer Research Institute, 61 62 Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Seoul, Republic of Korea, ¹¹American Cancer Society, Atlanta, GA, USA, ¹²Department of 63 64 Etiology & Carcinogenesis and State Key Laboratory of Molecular Oncology, Cancer Institute 65 and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, ¹³Department of Epidemiology, School of Public Health, China Medical 66 University, Shenyang, China, ¹⁴Department of Clinical Epidemiology and Center of Evidence 67 Based Medicine, The First Hospital of China Medical University, Shenyang, China, ¹⁵Ministry of 68 69 Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China, ¹⁶State Key Laboratory of Genetic Engineering, School of Life 70 Sciences, Fudan University, Shanghai, China, ¹⁷Saw Swee Hock School of Public Health, 71 72 National University of Singapore and National University Health System, Singapore, Singapore, 73 ¹⁸Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore 74 and National University Health System, Singapore, Singapore, ¹⁹National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan, ²⁰Department of Applied 75 Mathematics, Chung Yuan Christian University, Chong-Li, Taiwan, ²¹Department of Preventive 76 77 Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea, 78 ²²Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, 79 Republic of Kore, ²³Guangdong Lung Cancer Institute, Guangdong Provincial Key Laboratory of Translational Medicine in Lung Cancer, Guangdong Provincial People's Hospital (Guangdong 80 Academy of Medical Sciences), Southern Medical University, Guangzhou, China, ²⁴Department 81 of Pathology, Queen Mary Hospital, Hong Kong, Hong Kong, ²⁵Department of Biostatistics, 82 83 Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA, 84 ²⁶Department of Biostatistics, Bioinformatics & Biomathematics, Georgetown University Medical Center, Washington, DC, USA, ²⁷Department of Computational Biology, St. Jude 85 Children's Research Hospital, Memphis, TN, USA, ²⁸Saville Cancer Screening and Prevention 86 87 Center, Inova Schar Cancer Institute, Inova Health System, Fairfax, VA, USA, ²⁹Department of 88 Environmental Health, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea., ³⁰Epidemiology and Community Health Branch, National Heart Lung and 89 90 Blood Institute, Bethesda, MD, USA, ³¹Shanghai Pulmonary Hospital, Shanghai, China, 91 ³²Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo,

⁴⁸

Japan, ³³Lung and Esophageal Cancer Clinic, Chonnam National University Hwasun Hospital, 92 93 Hwasuneup, Republic of Korea, ³⁴Department of Internal Medicine, Chonnam National 94 University Medical School, Gwangju, Republic of Korea, ³⁵Cancer Institute and Hospital, 95 Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, ³⁶Department of Medicine, School of Clinical Medicine, The University of Hong Kong, Queen 96 97 Mary Hospital, Hong Kong, Hong Kong, ³⁷Laboratory for Genotyping Development, RIKEN 98 Center for Integrative Medical Sciences, Yokohama, Japan, ³⁸Center for Antibody and Vaccine 99 Therapy, Research Hospital, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, ³⁹Department of Medical Oncology and Cancer Center, and Center for Advanced 100 101 Medicine against Cancer, Shiga University of Medical Science, Shiga, Japan, ⁴⁰Department of 102 Statistics & Research Institute of Natural Sciences, Sookmyung Women's University, Seoul, 04310, Korea, ⁴¹Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical 103 104 Sciences, Yokohama, Japan, ⁴²Department of Thoracic Surgery, Tokyo Medical and Dental University, Tokyo, Japan, ⁴³Department of Respiratory Medicine, Tokyo Medical and Dental 105 University, Tokyo, Japan, ⁴⁴Department of Epidemiology and Population Health, Albert Einstein 106 107 College of Medicine, New York, USA, ⁴⁵Department of Medical Oncology, Japanese Red Cross Medical Center, Tokyo, Japan, ⁴⁶Department of Thoracic Surgery, National Cancer Center 108 Hospital, Tokyo, Japan, ⁴⁷Molecular Pathology and Genetics Division, Kanagawa Cancer Center 109 Research Institute, Yokohama, Japan, ⁴⁸Department of Thoracic Surgery, Kanagawa Cancer 110 111 Center, Yokohama, Japan, ⁴⁹Department of Thoracic Oncology, National Cancer Center Hospital East, Kashiwa, Japan, ⁵⁰Department of Thoracic Surgery, National Cancer Center Hospital East, 112 Kashiwa, Japan, ⁵¹Division of Medical AI Research and Development, National Cancer Center 113 114 Research Institute, Tokyo, Japan, ⁵²Department of Cellular and Organ Pathology, Graduate School of Medicine, Akita University, Akita, Japan, ⁵³Department of Thoracic Surgery, Graduate 115 School of Medicine, Akita University, Akita, Japan, ⁵⁴Department of Preventive Medicine, 116 Faculty of Medicine, Saga University, Saga, Japan, ⁵⁵Department of Preventive Medicine, 117 Nagoya University Graduate School of Medicine, Nagoya, Japan, ⁵⁶Laboratory of Clinical 118 Genome Sequencing, Department of Computational Biology and Medical Science. Graduate 119 120 School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, ⁵⁷Department of Molecular 121 Biology, Institute for Advanced Medical Sciences, Nippon Medical School, Tokyo, Japan, ⁵⁸Department of Surgery, Division of General Thoracic Surgery, Shinshu University School of 122 Medicine Asahi, Nagano, Japan, ⁵⁹Department of Chest Surgery, Fukushima Medical University 123 School of Medicine, Fukushima, Japan, ⁶⁰Department of Gastrointestinal Tract Surgery, 124 125 Fukushima Medical University School of Medicine, Fukushima, Japan, ⁶¹Department of Integrative center of General Surgery, Gunma University Hospital, Gunma, Japan, ⁶²Institute of 126 Occupational Medicine and Ministry of Education Key Lab for Environment and Health, School 127 of Public Health, Huazhong University of Science and Technology, Wuhan, China, ⁶³China 128 129 National Environmental Monitoring Center, Beijing, China, ⁶⁴Department of Internal Medicine, 130 Division of Oncology/Hematology, College of Medicine, Korea University Anam Hospital, Seoul. Republic of Korea, ⁶⁵Department of Clinical Oncology, The University of Hong Kong, 131 Queen Mary Hospital, Hong Kong, Hong Kong, ⁶⁶School of Medicine and Institute of Medicine, 132 Chung Shan Medical University, Taichung, Taiwan, ⁶⁷Department of Internal Medicine, Division 133 of Pulmonary Medicine, Chung Shan Medical University Hospital, Taichung, Taiwan, ⁶⁸Institute 134 of Biomedical Sciences, National Chung Hsing University, Taichung, Taiwan, ⁶⁹Department of 135 136 Internal Medicine, Division of Chest Medicine, Taichung Veterans General Hospital, Taichung, 137 Taiwan, ⁷⁰Department of Respiratory Therapy, Chang Gung University, Taoyuan, Taiwan,

⁷¹Department of Pulmonary and Critical Care, Xiamen Chang Gung Hospital, Xiamen, China, 138 139 ⁷²Department of Internal Medicine, National Taiwan University Hospital and College of 140 Medicine, Taipei, Taiwan, ⁷³Department of Internal Medicine, E-Da Cancer Hospital, I-Shou 141 University and Kaohsiung Medical University, Kaohsiung, Taiwan, ⁷⁴Department of Oncology, 142 National Cheng Kung University Hospital, College of Medicine, National Cheng Kung 143 University, Tainan, Taiwan, ⁷⁵Department of Chest Medicine, Taipei Veterans General Hospital, 144 and school of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ⁷⁶Lung 145 Cancer Center, Kyungpook National University Medical Center, Daegu, Republic of Korea, ⁷⁷Jeonnam Regional Cancer Center, Chonnam National University, Hwasun, Republic of Korea, 146 147 ⁷⁸Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China, ⁷⁹Department of Human Genetics, Genome Institute of Singapore, Singapore, Singapore, ⁸⁰School of Life 148 Sciences, Anhui Medical University, Hefei, China, ⁸¹Karmanos Cancer Institute, Wayne State 149 150 University, Detroit, MI, USA, ⁸²Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK, ⁸³Department of Medicine, Section of Epidemiology and Population 151 Science, Institute for Clinical and Translational Research, Houston, TX, USA, ⁸⁴School of 152 153 Medicine, Zhejiang University, Hangzhou, Zhejiang, China, ⁸⁵Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, USA, ⁸⁶Department of Integrative Oncology, BC 154 Cancer, Vancouver, BC, Canada, 87 Health Research Institute of Asturias (ISPA), Oviedo and 155 Universidad Nebrija, Madrid, Spain, ⁸⁸Program in Epidemiology, Public Health Sciences 156 157 Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA, ⁸⁹Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, ⁹⁰Department of Clinical 158 159 Biochemistry, Herley and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark, ⁹¹International Agency for Research on Cancer (IARC/WHO), Lyon, France, 160 ⁹²German Cancer Research Center (DKFZ), Heidelberg, Germany, ⁹³Translational Lung 161 162 Research Center Heidelberg (TLRC-H), Member of the German Center for Lung Research 163 (DZL), Heidelberg, Germany, ⁹⁴Department of Biosciences and Medical Biology and Center for 164 Tumorbiology and Immunology (CTBI), University of Salzburg and Cancer Cluster Salzburg, Salzburg, Austria, ⁹⁵University Medical Center Goettingen, Goettingen, Germany, ⁹⁶Institute of 165 166 Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany, ⁹⁷Helmholtz Center Munich, Institute of Epidemiology, Munich, Germany, ⁹⁸Institute 167 of Medical Statistics and Epidemiology, Technical University Munich, Munich, Germany, 168 ⁹⁹Harvard TH Chan School of Public Health, Boston, Massachusetts, USA, ¹⁰⁰Carmel Medical 169 Center, Israel, ¹⁰¹Markey Cancer Center, Lexington, KY, USA, ¹⁰²Liverpool University, 170 171 Liverpool, UK, ¹⁰³The University of Texas MD Anderson Cancer Center, Houston, Texas, USA, 172 ¹⁰⁴Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA, ¹⁰⁵Princess Margaret Cancer Center, Toronto, ON, Canada, ¹⁰⁶Norris Cotton Cancer Center, Lebanon, NH, 173 USA, ¹⁰⁷Radboud University Medical Center, Nijmegen, Netherlands, ¹⁰⁸National Institute of 174 175 Occupational Health, Oslo, Norway, ¹⁰⁹Department of Medical Biosciences, Umeå University, 176 Umeå, Sweden, ¹¹⁰Department of Radiation Sciences, Umeå University, Umeå, Sweden, ¹¹¹University of Sheffield, Sheffield, UK, ¹¹²UPMC Hillman Cancer Center and Department of 177 Epidemiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA, ¹¹³D, 178 179 School of Pharmacy and Pharmaceutical Sciences, SUNY University at Buffalo, NY, USA, ¹¹⁴Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, 180 Tampa, FL, USA, ¹¹⁵Department of Thoracic Surgery, Division of Epidemiology, Vanderbilt 181 University Medical Center, Nashville, TN, USA, ¹¹⁶Cancer Research Center, Kyungpook 182 National University Medical Center, Daegu, Republic of Korea, ¹¹⁷Department of Thoracic and 183

184 Cardiovascular Surgery, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea, ¹¹⁸Department of Occupational and Environmental Health 185 186 and Ministry of Education Key Lab for Environment and Health, School of Public Health, Tongji 187 Medical College, Huazhong University of Science and Technology, Wuhan, China, ¹¹⁹Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos 188 Biomedical Research Inc, Rockville, MD, USA, ¹²⁰Division of Applied Oral Sciences & 189 190 Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, 191 China, ¹²¹State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, 192 Hong Kong, Chinaboratory of Pharmaceutical Biotechnology, The University of Hong Kong, 193 Hong Kong, China, ¹²²State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, China, ¹²³Department of Oncology, Cancer Center, Union Hospital, 194 195 Huazhong University of Science and Technology, Wuhan, China, ¹²⁴The Ph.D. Program for 196 Translational Medicine, College of Medical Science and Technology, Taipei Medical University, 197 Taipei, Taiwan, ¹²⁵School of Public Health, Li Ka Shing (LKS) Faculty of Medicine, The 198 University of Hong Kong, Hong Kong, China, ¹²⁶Key Laboratory of Cancer Etiology and 199 Intervention, University of Liaoning Province, Shenyang, China, ¹²⁷Department of Internal 200 Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan, ¹²⁸Gleneagles Hong Kong Hospital, Hong Kong, China, ¹²⁹Department of Internal Medicine, 201 Division of Medical Oncology, College of Medicine, Korea University Guro Hospital, Seoul, 202 203 Republic of Korea, ¹³⁰Department of Internal Medicine, Seoul National University Bundang 204 Hospital, Seongnam, Republic of Korea, ¹³¹Department of Oncology, Wuhan Iron and Steel 205 (Group) Corporation Staff-Worker Hospital, Wuhan, China, ¹³²Division of Pulmonary Oncology 206 and Interventional Bronchoscopy, Department of Thoracic Medicine, Chang Gung Memorial 207 Hospital, Taoyuan, Taiwan, ¹³³Division of Thoracic Surgery, Department of Surgery, Chung Shan Medical University Hospital, Taichung, Taiwan, ¹³⁴Division of Environmental 208 Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The 209 Netherlands, ¹³⁵Department of Epidemiology and Biostatistics, University of California, San 210 211 Francisco, San Francisco, CA, USA, ¹³⁶Department of Pathology, Queen Elizabeth Hospital, 212 Hong Kong, China, ¹³⁷Qujing Center for Diseases Control and Prevention, Qujing, China, 213 ¹³⁸Department of Oncology, Shandong Cancer Hospital and Institute, Shandong Academy of Medical Sciences, Jinan, China, ¹³⁹Division of Cohort Research, National Cancer Center Institute 214 for Cancer Control, Tokyo, Japan, ¹⁴⁰Division of Epidemiology, National Cancer Center Institute 215 for Cancer Control, Tokyo, Japan, ¹⁴¹Center for Global Health, National Cancer Institute, 216 217 Bethesda, MD, USA, ¹⁴²Institute of Medicine, Chung Shan Medical University, Taichung, 218 Taiwan, ¹⁴³Genomic Research Center, Academia Sinica, Taipei, Taiwan, ¹⁴⁴Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan, ¹⁴⁵Laboratory Services, 219 220 American Cancer Society, Georgia, USA, ¹⁴⁶Department of Oncology, School of Medicine, 221 Johns Hopkins University, Baltimore, MD, USA, ¹⁴⁷Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, ¹⁴⁸Institute for Clinical and 222 Translational Research, Baylor College of Medicine, Houston, TX, USA, ¹⁴⁹Section of 223 Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, 224 225 Houston, TX, USA, ¹⁵⁰Dan L Duncan Comprehensive Cancer Center, Baylor College of 226 Medicine, Houston, TX, USA

227

^{*} These authors contributed equally to this work

[‡] These authors jointly directed this work

230

- 231 Correspondence to: Batel Blechter (<u>batel.blechter@nih.gov</u>), Qing Lan (<u>qingl@mail.nih.gov</u>),
- and Haoyu Zhang (<u>haoyu.zhang2@nih.gov</u>)

233

234

235

236

238 Abstract

239

240	Polygenic risk scores (PRSs) are	promising	for risk	stratification	but have mainly	v been
				101 11011		000000000000000000000000000000000000000	,

- 241 developed in European populations. This study developed single- and multi-ancestry PRSs for
- 242 lung adenocarcinoma (LUAD) in East Asian (EAS) never-smokers using genome-wide
- association study summary statistics from EAS (8,002 cases; 20,782 controls) and European
- 244 (2,058 cases; 5,575 controls) populations. A multi-ancestry PRS, developed using CT-SLEB,
- 245 was strongly associated with LUAD risk (odds ratio=1.71, 95% confidence interval
- 246 (CI):1.61,1.82), with an area under the receiver operating curve value of 0.640 (95%
- 247 CI:0.629,0.653). Individuals in the highest 20% of the PRS had nearly four times the risk
- compared to the lowest 20%. Individuals in the 95th percentile of the PRS had an estimated
- 249 6.69% lifetime absolute risk. Notably, this group reached the average population 10-year LUAD
- risk at age 50 (0.42%) by age 41. Our study underscores the potential of multi-ancestry PRS

approaches to enhance LUAD risk stratification in EAS never-smokers.

- 252
- 253
- 254
- 255
- 256
- 257
- 258
- 259
- 260
- 261
- 262
- 263
- 264
- 265
- 266
- 267

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

271 Abbreviations:

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

291 Introduction

292

Lung cancer is a major global health challenge, responsible for about 20% of all cancer deaths in 2020¹. While smoking is the primary etiologic factor, around 25% of lung cancer cases occur in never-smokers, with significant geographical variations². Notably, in East Asia (EAS), never-

smoking women exhibit high incidences of lung adenocarcinoma (LUAD), the predominant

- 297 histologic subtype³.
- 298

299 Beyond smoking, risk factors for LUAD include environmental and occupational exposures,

300 lifestyle, family history, and genetic susceptibility^{4–7}. Genome-wide association studies (GWAS)

301 for never-smoking lung cancer across EAS and European (EUR) populations have identified 28

302 susceptibility loci at 25 independent regions^{8–13}. These GWAS findings enhance our

303 understanding of LUAD's genetic architecture, yet the translation of these findings into clinical

304 practice requires further investigation.

305

306 Polygenic risk scores (PRS) aggregate the effects of individual single nucleotide polymorphisms 307 (SNPs) to estimate genetic risk for diseases, playing a crucial role in cancer prevention and 308 screening. Previous efforts primarily utilized EUR populations for lung cancer PRS development^{14,15} to better identify high-risk individuals. Nonetheless such PRSs often fall short 309 310 for EAS never-smokers, reflecting a bias in screening guidelines and the focus of studies on 311 populations with EUR ancestry. Despite ongoing efforts to integrate PRS into clinical practices 312 for complex diseases, as seen in initiatives like the electronic Medical Records and Genomics 313 (eMERGE) network¹⁶, Veterans Affairs Genomic Medicine at Veteran Affairs (GenoVA) 314 study¹⁷, and the Women Informed to Screen Depending On Measures of risk (WISDOM) 315 study¹⁸, lung cancer has been excluded. The predominant global burden of lung cancer, driven 316 primarily by smoking, overlooks the significant impact and potential benefits of PRS in never-317 smoking populations.

318

319 The recent expansion of GWAS across diverse populations, combined with advancements in

320 PRS methodology, underscores a pivotal shift towards enhancing disease outcome prediction

321 beyond EUR populations¹⁹. Notably, the development of multi-ancestry PRS methods marks a

considerable advancement, leveraging data from various ancestral backgrounds to enrich the
 predictive accuracy and robustness of PRSs^{20–26}. By jointly modeling genetic data from multiple
 populations, these approaches improve the predictive power of PRSs, thereby enhancing disease
 outcome predictions in non-EUR populations.

326

327 While the development and validation of PRSs are critical, the application of an established PRS

in estimating the absolute risk of a disease offers valuable insights into risk stratification,

329 potentially guiding clinical interventions, such as lung cancer screening. Moreover, projecting

ancessary sample sizes for future studies to achieve varying levels of PRS predictive accuracy is

an underexplored area that could significantly impact epidemiological research design and cost-

332 effectiveness.

333

This study aims to evaluate the predictive performance of single- and multi-ancestry PRSs for

335 LUAD in never-smoking EAS individuals using state-of-the-art methodologies. By integrating

multiple datasets, we constructed PRSs based on summary statistics for 5,622 never-smoking

cases and 21,813 never-smoking controls from EAS and EUR ancestries. We assessed the

performance of the PRSs using independent, individual level data of 4,438 never-smoking EAS

cases and 4,544 never-smoking EAS controls. Furthermore, we estimated the lifetime and 10-

340 year absolute risks of LUAD using the most accurate PRSs developed. Lastly, we projected the

341 sample sizes needed in future research to achieve specific levels of prediction accuracy with

342 PRSs in EAS never-smokers (Figure 1).

343

345	Methods
346	
347	Study design and data sources
348	
349	Training data compilation: study population and genotyping
350	The studies, genotyping protocols, and quality control for PRS construction are previously
351	described in detail ²⁷ . Briefly, EAS single-ancestry PRSs for LUAD were constructed using
352	summary data from 3,564 never-smoking LUAD cases and 16,238 never-smoking controls of
353	EAS ancestry from the Nanjing Lung Cancer Study (NJLCS) ^{9,28} , National Cancer Center of
354	JAPAN (NCC), and the Research Institute and Aichi Cancer Center (ACC). The NJLCS study
355	combined data from several cities, genotyped by Affymetrix Genome-Wide Human SNP Array
356	6.0, as well as the Oncoarray GWAS ²⁸⁻³⁰ . The NCC study included lung cancer patients and
357	control data from multiple Japanese, with genotyping conducted on Illumina
358	HumanOmniExpress and HumanOmni1-Quad genotyping platforms. Similarly, the ACC study
359	pooled data from the multiple Japanese medical institutions, and the Nagahama Study, with
360	genotyping performed using Illumina 610k and Illumina660k platforms. To construct the multi-
361	ancestry PRS, we further incorporated GWAS summary statistics from 2,058 never-smoking
362	LUAD cases and 5,575 never-smoking controls of EUR ancestry, genotyped using Illumina
363	Infinium OmniExpress-24 v1.2 BeadChips and Illumina Human660W-Quad BeadChip ¹³ .
364	
365	Tuning and validation: study population and genotyping
366	For PRS tuning and validation, we used 4,438 never-smoking LUAD cases and 4,544 never-
367	smoking controls from the Female Lung Cancer Consortium in Asia (FLCCA) ^{10,11} , an
368	international consortium composed of never-smoking EAS women from regions including
369	Mainland China, Singapore, Taiwan, Hong Kong, South Korea and Japan ^{10,11} . All samples were
370	genotyped using the Illumina 660W, 370K or 610Q microarrays. The FLCCA data were
371	randomly and equally divided for tuning (2,219 cases, 2,272 controls) and validation (2,219
372	cases, 2,272 controls). Informed consent was obtained from all participants, and the recruitment
373	and data collection procedures were approved by the local ethics review committees of all
374	participating research institutes.

376 Construction of PRS

377

378 Single-ancestry PRS methods

379 **PRS-25**: We constructed a simple PRS using the 25 independent SNPs (PRS-25) that reached

380 genome-wide significance (i.e., $P < 5x10^{-8}$) in the largest GWAS for LUAD in EAS population to

date⁸. The PRS-25 was calculated by summing the risk alleles weighed by their effect sizes (i.e.,

382 per allele log-odds ratio) obtained from a meta-analysis including only never-smokers.

383

384 **PRS-CT:** We generated a PRS using the CT method³¹ using PLINK 1.90³². This involved the

clumping of SNPs based on linkage disequilibrium (LD) with a threshold of $r^2=0.1$ within a

386 500kb window. The LD were estimated from half of the FLCCA dataset used for tuning dataset.

387 We created nine SNP subsets by applying incremental P-value thresholds (i.e., $P < 5x10^{-8}$,

388 P<5x10⁻⁷,..., P<5x10⁻¹, 1), and calculated their respective PRSs using PLINK2³² command "--

389 score cols=+scoresums,-scoreavgs no-mean-imputation". Using the tuning dataset, the AUC was

390 calculated for each threshold to identify the most predictive P-value cutoff.

391

392 LDpred2-PRS: We applied the LDpred2 method, implemented in the R package bigsnpr³³, 393 using a Bayesian framework to estimate SNP effect sizes through a shrinkage estimator. This 394 method leverages GWAS summary statistics, incorporating a prior for effect sizes, while also 395 accounting for LD across SNPs. Our analysis was constrained to HapMap3 variants, and we 396 calibrated the model using a range of hyperparameters: the proportion of causal SNPs was set across a 21-point logarithmic sequence from 10^{-5} to 1, and the per-SNP heritability was set as a 397 398 fraction (0.3, 0.7, 1 or 1.4) of the total heritability estimated by LD score regression³⁴ divided by 399 the number of causal SNPs. Lastly, we used the "sparse" option to set weaker effects to zero. The 400 optimal tuning parameters were selected based on the highest AUC achieved on the tuning 401 dataset.

402

403 PRS-EUR 128: We evaluated the performance of a European-derived PRS (PRS-EUR 128)³⁵,
404 which consists of 128 variants obtained from a GWAS of predominately active smokers, on EAS
405 populations to assess its cross-population applicability.

407 <u>Multi-ancestry PRSs methods:</u>

408 **LDpred2 PRS + PRS-EUR 128**: For the multi-ancestry PRSs, we applied the weighted-PRS

- 409 approach by linearly combining the most predictive single-ancestry PRS from the EAS
- 410 population (Ldpred2 PRS) with the EUR population-specific PRS (PRS-EUR 128). The weights
- 411 of this combined PRS were calculated by applying a logistic regression on the tuning dataset in R
- 412 version 4.2.0.
- 413

414 **PRS-CSx**: We applied PRS-CSx, a multi-ancestry polygenic prediction method that uses a

415 Bayesian framework with a continuous shrinkage prior to estimate SNP effect sizes from GWAS

- 416 summary statistics across different populations²⁶. LD reference panels for EUR and EAS,
- 417 provided by the PRS-CSx software, were constructed from the 1000 Genomes Project (1KG)
- 418 samples ³⁶. We used the default setting with the gamma-gamma prior hyperparameters (a and b)
- 419 at 1 and 0.5, respectively. The shrinkage parameter ϕ was assessed at 10⁻⁶, 10⁻⁴, 10⁻², and 1
- 420 within the tuning dataset to select the value with highest AUC. With the optimal ϕ , we
- 421 calculated the PRS using posterior effect-sizes from both EAS and EUR populations. Weights
- 422 were then estimated to linearly combine the two PRSs on the tuning dataset. The final
- 423 performance was then evaluated on an EAS independent validation dataset.
- 424
- 425 **CT-SLEB:** We used the recently developed CT-SLEB method to derive a PRS using EAS and
 426 EUR datasets ²⁴. This method extends the standard CT method by designing a two-dimensional
 427 approach to select SNPs for EAS PRS construction. It incorporates an empirical Bayesian (EB)
 428 framework to model genetic correlations between EAS and EUR. Following this, a super429 learning (SL) model is then applied to integrate multiple PRSs, each generated using distinct p-
- 429 learning (SL) model is then applied to integrate multiple PRSs, each generated using distinct p
- 430 value thresholds and clumping parameters³⁷. Our implementation of CT-SLEB followed the
- 431 default setting with p-value thresholds $p_t = 5 \times 10^{-8}$, 5×10^{-7} , ..., 5×10^{-1} , or 1, and genetic

432 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

- 433 process for EAS was based on the LD reference derived from the FLCCA tuning dataset, and the
- 434 EUR LD reference was based on the 1KG EUR samples.
- 435
- 436 Relative and absolute risk calculation for PRSs
- 437

438 To evaluate each PRS's performance, we standardized the scores to a unit standard deviation, 439 based on the distribution among the control group in the EAS validation dataset. We then 440 quantified the association between the standardized PRSs and LUAD risk using the OR and 95% 441 CI, via logistic regression, adjusting for age and the first 10 principal components (PCs). We 442 evaluated the predictive performance of the different PRS models through the adjusted AUC 443 values, accounting for age and top 10 PCs, using the R package RISCA³⁷. We further estimated 444 the ORs of each PRS for risk of LUAD based on PRS percentiles, setting the middle quantile 445 (40-60%) as the reference category. Lastly, we used the likelihood ratio test to evaluate the interaction between the PRS and age ($<40, 40-49, 50-59, 60-69, \geq 70$). 446 447 For absolute risk estimates, we used the iCARE software³⁸ to calculate the cumulative lifetime 448 449 (age 30-80) and 10-year absolute risks of LUAD among never-smoking female controls in 450 FLCCA (N=4,544). Absolute risks were derived by applying the Cox proportional hazard model 451 with the top performing PRS (CT-SLEB) and age-specific lung cancer incidence and mortality rates in Taiwan³⁹. We estimated the absolute risk of LUAD overall, as well as stratified by first-452

- 453 degree family history of lung cancer.
- 454

455 Projections of CT and LDpred2 PRS performance by sample size

456

457 We used GENESIS package ⁴⁰ to estimate the sample sizes needed for PRS to reach various 458 AUC levels in the EAS population. This method estimates the expected number of SNP 459 discoveries and their explained heritability in future studies. Using the GWAS summary statistics 460 from our EAS training dataset, and the provided LD scores for EAS populations from the 1KG dataset, we projected the AUC for CT PRS across various case-control ratios, from 1:1 to 1:10. 461 462 and case numbers ranging from 5,000 to 200,000. Given GENESIS's specific design for CT PRS, 463 we developed a extend its application to LDpred2 PRS projections. This involved modeling the 464 relationship between effective sample sizes with the phenotypic variance ratio between LDpred2 465 and CT PRS (Supplementary Figure 1), elaborated in the Supplementary Note. 466

- 467 **Results**
- 468

469 Development and validation of the PRS

- 470
- 471 We applied several cutting-edge single-and multi-ancestry PRS methods (Methods,
- 472 Supplementary Figure 2), evaluating their performance in terms of relative risk and AUC
- 473 within the EAS validation dataset (Methods, Table 1, Supplementary Figure 2-3). Among
- 474 single-ancestry PRS methods, the LDpred2 PRS, incorporating 942,591 SNPs, outperformed
- 475 other methods, with an odds ratio (OR) per unit standard deviation (SD) of 1.62 (95% confidence
- 476 interval (CI): 1.52, 1.73) alongside an adjusted AUC of 0.629 (95% CI: 0.618, 0.641). In
- 477 contrast, the PRS designed exclusively for EUR populations underperformed in our EAS
- 478 population, with an estimated adjusted AUC of 0.489 (95% CI: 0.477, 0.501), possibly attributed
- to the inclusion of smokers in the development of the existing EUR PRS.
- 480
- 481 Among multi-ancestry PRS methods, the weighted-PRS of EAS and EUR yielded an OR per unit
- 482 SD of 1.62 (95% CI: 1.52, 1.72), with PRS-CSx and CT-SLEB showing even stronger
- 483 association with ORs of 1.67 (95% CI: 1.57, 1.78) and 1.71 (95% CI: 1.61, 1.82), respectively.
- 484 Among all the PRS methods, CT-SLEB had the highest AUC of 0.640 (95% CI: 0.629, 0.653).
- 485 Setting individuals in the middle PRS quantile (40th to 60th percentile) as the reference category,
- those in the highest 5% of risk for the top performing PRS, CT-SLEB, had 4.17 (95% CI: 3.20,
- 487 5.47)-fold risk, whereas those in the lowest 5% had 0.33 (95% CI: 0.21, 0.51)-fold risk of
- 488 developing LUAD (**Figure 2**).
- 489
- 490 Interaction between CT-SLEB PRS and age
- 491

492 We observed a significant multiplicative interaction between the CT-SLEB PRS and age at 493 diagnosis for LUAD risk (p-interaction = 0.002). As illustrated in **Figure 3**, the OR per SD of 494 PRS was higher in younger individuals, particularly those under 40 years of age (OR=1.70, 95% 495 CI: 1.42, 2.05) compared to individuals aged 70 and older (OR=1.53, 95% CI: 1.37, 1.72). The 496 association followed a roughly linear pattern across middle-age categories, peaking between ages 497 60-69 (OR = 1.93, 95% CI: 1.79-2.08). This pattern suggests a stronger impact of genetic risk 498 factors in younger age groups. A similar trend was observed when analyzing the association 499 between LUAD risk and a dichotomized PRS (upper 90th percentile) by age categories.

500

501 Absolute risk of developing lung adenocarcinoma by CT-SLEB PRS

502

503 Using the iCARE package⁴¹, we estimated the absolute risk of LUAD for never-smoking EAS 504 women, utilizing the CT-SLEB PRS which demonstrated the highest AUC (Table 1). The 505 cumulative lifetime risk of LUAD, assessed between age 30-80, varied markedly across PRS 506 percentiles, ranging from a minimal mean of 0.78% in the 5th percentile to a substantial mean of 507 6.69% in the 95th percentile (Figure 4). Additionally, the 10-year absolute risk for LUAD in 508 never-smoking 50-year-old women, varied from a mean of 0.13% in the 5th percentile to mean 509 of 1.11% in the 95th percentile of the PRS distribution. While the average 10-year absolute risk 510 for this age group is 0.42%, never-smoking women in the >95 and 90-95th percentiles PRS 511 reached this risk earlier, at age of 41 and 42, respectively. 512 513 To evaluate the combined effect of PRS and family history, we further modeled the absolute risk 514 for LUAD in individuals with (n=86) and without (n=1,414) family history of lung cancer in a 515 first-degree relative (Figure 5, Supplementary Figure 4). We observed that those with family 516 history of lung cancer have higher absolute lifetime risk of LUAD than those without family history. The lifetime absolute risk between age 30-80 for women at the bottom 5th and top 95th 517 518 percentile of the PRS with family history of lung cancer were 2.77% and 9.39%, respectively 519 (Supplementary Figure 4). Additionally, the 10-year absolute risk for women in the 95th PRS 520 percentile with a family history reached the population average 10-year absolute risk at age 50 521 (0.42%) much earlier, at age 38. Formal tests for interaction between PRS and family history 522 were not statistically significant (p=0.45), suggesting that family history and PRS independently 523 contribute to LUAD risk stratification.

524

525 Projections of polygenic risk score performance by sample size

526

527 Using the GENESIS model, we projected the expected AUC of PRS-CT and LDpred2 PRS

528 under varying GWAS sample sizes and case-control ratios for LUAD in never smoking EAS

529 women (Figure 6). Our analysis, using the EAS training dataset, estimated approximately 1,772

530 (s.e.=1,641) susceptibility variants that are independently associated with LUAD²⁷. This high

531	number of variants underscores the extensive polygenic nature of LUAD, implying relatively
532	small effect sizes for individual SNPs. Based on our current data, the expected AUC for the
533	LDpred2 PRS stands at 0.631 (95% CI: 0.618, 0.641), aligning closely with the actual
534	performance of the LDpred2 PRS in the validation dataset. The upper boundary of our
535	predictions, based solely on PRS, suggests an AUC potential of up to 0.731 (95% CI:
536	0.640,0.786). As sample sizes increase in our projections, the AUC is estimated to rise to 0.673
537	(95% CI: 0.593–0.725) with 55,000 EAS LUAD cases and a 1:1 case-control ratio. With the
538	same number of cases but a 1:10 case-control ratio, the AUC is projected to further increase to
539	0.698 (95% CI: 0.621–0.727), nearing the theoretical upper limit of predictive accuracy.
540 541	
542	Explaining genetic variance through PRS across different sample sizes
543	
544	We also evaluated the genetic variance explained by the PRS under different sample sizes given
545	the projected results (Supplementary Note, Table 2). The current LDpred2 PRS explained
546	26.6% of the genome-wide chip heritability, contributing to approximately 16.5% of the 1.84-
547	fold familial relative risk for lung cancer among EAS never-smokers. Concurrently, genome-
548	wide chip variants heritability explained 61.9% of the 1.84-fold familial relative risk associated
549	with the disease. With an expanded sample size of 35,000 cases and 350,000 controls, the
550	constructed PRS is projected to account for 57.9% of the genome-wide chip variants heritability,
551	and 35.8% of the 1.84-fold familial relative risk for the disease.
552	
553	Discussion
554	
555	We developed and validated single- and multi-ancestry PRSs for LUAD in never-smoking EAS
556	individuals using the largest GWAS dataset of never-smokers to date. The multi-ancestry PRS
557	method, CT-SLEB, integrating summary data from EAS and EUR never-smokers, emerged as
558	the best-performing PRS. It exhibited a dose-response relationship with LUAD risk and achieved
559	higher AUC than all other evaluated PRSs. Further, our analysis demonstrates the potential of

560 PRS in stratifying individuals' 10-year and lifetime risk of developing LUAD. Lastly, we

561 projected the expected discriminatory accuracy of the PRS across a range of sample sizes and 562 case-control ratios.

563

To date, GWAS and subsequent PRS models have largely centered on EUR populations ^{42,43}, 564 565 rendering them less precise when applied to non-EUR populations and risking the exacerbation of health disparities ^{44,45}. Consistent with prior research ¹⁵, our study observed that a lung cancer 566 567 PRS constructed using EUR data, which included both smokers and never-smokers, significantly 568 underperformed when applied to EAS never-smoking individuals (AUC = 0.489, 95% CI: 569 (0.477, 0.501)), further highlighting the heterogeneity of the disease across populations and the 570 need to expand risk assessment efforts to non-EUR populations. Notably, by utilizing novel 571 methods to integrate GWAS data from both EUR and EAS populations, we demonstrated that a 572 multi-ancestry PRS enhances the precision of risk stratification for LUAD among EAS never-573 smokers. 574 575 Large-scale efforts to integrate PRS into clinical practice have focused on conditions with wellestablished PRS prediction performance^{16–18}, such as Type 2 Diabetes, breast cancer, and 576 577 cardiovascular diseases. Lung cancer has been notably absent from these efforts, primarily due to 578 the focus on smoking as a risk factor. This has limited the exploration and application of PRS in 579 lung cancer risk stratification, particularly among never-smokers. Recent efforts, such as the TALENT study in Taiwan⁴⁶, have evaluated low-dose computed tomography (LDCT) screening 580 581 for never-smokers with traditional risk factors like family history, passive smoking exposure, and 582 chronic lung diseases. However, a PRS has not yet been integrated into these screening 583 protocols. Further, a recent study in Taiwan incorporating genetic susceptibility in a risk model

584 with lifestyle and environmental risk factors observed an improvement in the AUC from 0.697 to

585 0.714³⁹. Notably, the study found that even by incorporating only 11 susceptibility loci, the

586 model's discriminative power had improved. Our study demonstrates the potential of a genome-

587 wide PRS to complement these efforts, enhancing risk stratification and identifying high-risk 588

individuals at younger ages, who may benefit from earlier interventions.

589

590 We address this gap by presenting a PRS that signifies a 3.92-fold increase in lung cancer risk 591 for individuals in the top 20% risk quantile, a marked improvement over the 2.09-fold increase

observed in a prior study by Wei et al.⁴⁷ for a cohort of Chinese never-smoking women. Further, 592 593 we identified a significant interaction between the PRS and age, observing heterogeneity in 594 effects of PRS across age categories. Similar to findings in a study on prostate cancer⁴⁸, we 595 observed higher associations between the PRS and LUAD among younger individuals compared 596 to those diagnosed at a later age. While the pattern was not strictly linear, and competing risks of 597 mortality may influence these results, our findings suggest that the relative impact of genetic 598 susceptibility may decrease as non-genetic risk factors accumulate with age. This highlights the 599 potential utility of PRS in younger populations, where genetic risk may play a more pronounced 600 role in disease onset.

601

602 We estimated the lifetime and 10-year absolute risk of LUAD using the PRS, which can be used 603 in risk stratification efforts and to identify those who are at high risk for the disease. 604 Specifically, we found that the 10-year absolute risk for LUAD in a never-smoking 50-year-old 605 woman, a critical age for initiating recommended annual lung cancer screening as per the United States Preventive Services Taskforce (USPSTF)⁴⁹, varied from 0.13% in the 5th percentile to 606 607 1.11% in the 95th percentile of the PRS distribution. Notably, women in the highest risk 608 categories (>95th and 90-95th percentiles) reached the average 10-year absolute risk (0.42%) 609 much earlier, at ages 41 and 42, respectively. This finding suggests that the PRS has the potential 610 to identify younger women who may benefit from earlier risk-based interventions. While limited 611 in sample size, we also found that the lifetime absolute risk of LUAD was even higher among 612 women in the top percentile of the PRS and with first-degree family history of lung cancer, 613 suggesting the two risk factors may act independently for their effect on LUAD risk. Although 614 the absolute risk values are modest, the PRS's ability to stratify women into higher-risk 615 categories at younger ages suggests potential utility in guiding early screening decisions. Early 616 identification through PRS could be particularly valuable for never-smoking individuals, who are 617 traditionally excluded from lung cancer screening programs. 618 619 The CT-SLEB method's improved performance can be attributed to its ability to utilize diverse

620 genotyping arrays, thereby including population-specific variants. This aspect, along with the

621 inclusion of EUR samples in the training set, broadens the genetic diversity and leverages larger

622 sample sizes for improved model accuracy. Moreover, CT-SLEB excels in modeling genetic

623 correlations across ancestries, efficiently using both shared and unique genetic markers for
624 refined disease risk estimations across populations. Lastly, we set up our study with a robust
625 three-sample split design, dividing datasets into training, tuning, and validation. This ensures that

- 626 PRS efficacy is validated independently, effectively reducing the overfitting risk.
- 627

Current PRS studies predominantly focus on evaluating relative and absolute risks^{8,12,15,47,50}, yet 628 629 our research introduces a crucial forecast of sample sizes required for differing levels of PRS 630 accuracy. Our projection analyses indicated that achieving an AUC of 0.70 would approach the 631 maximized prediction potential of the PRS, given the estimated genetic variance explained by 632 GWAS chip variants. To reach this level of accuracy, a future study would need to include 633 55,000 cases with a 1:10 case-control ratio. Notably, accessing large biobanks with publicly available controls could reduce the number of required cases⁵¹.-These projections offer a 634 635 strategic framework for planning and designing future genetic studies on lung cancer, 636 establishing clear benchmarks for PRS model performance, which is essential for the scientific

- 637 community in advancing genetic risk prediction.
- 638

639 Our study has several limitations. First, it focuses on the genetic susceptibility of LUAD without 640 considering additional questionnaire-based risk factors. However, to be able to utilize data from 641 the largest GWAS conducted to date, this work aimed primarily to identify the most effective 642 PRS model within genetic contexts that can then be used in future efforts. Further, we plan to 643 conduct integrative analyses that merge PRS with other risk indicators to refine LUAD risk 644 predictions for never-smokers. Second, our projections are tailored to single-ancestry PRS 645 models, not fully addressing the intricacies of multi-ancestry approaches, including diverse 646 sample sizes and genetic correlations across populations. Third, our validation cohort, derived 647 from the FLCCA and covering various EAS regions, relied on Taiwanese incidence data for 648 absolute risk estimations. Consequently, our absolute risk findings may not extend universally to 649 all EAS never-smokers, though our PRS performance assessment and sample size projections 650 remain applicable to our population of interest. Lastly, the clinical application of our PRS findings, particularly in screening and risk counseling, is still in preliminary stages. Despite 651 advancements in using PRS for smoking cessation trials⁵², its implementation for never-smokers 652 653 in clinical settings has been limited, presenting a significant area for future research.

65	54
0.	/ T

655	In summary, our study evaluates various PRS models to capture the genetic predisposition to
656	LUAD among never-smoking EAS individuals. It extends beyond risk prediction by estimating
657	both 10-year and cumulative lifetime absolute risks, and by projecting the sample sizes required
658	for future GWAS to refine the predictive power of PRSs in future GWAS. Additionally, we
659	quantify the phenotypic variance captured by PRSs across different sample sizes. Future studies
660	are crucial to further improve these PRS models, aiming to enhance genetic risk predictions
661	while integrating a wider array of risk factors. Such efforts will develop more accurate and
662	comprehensive risk models for LUAD in never-smoking individuals across diverse populations.
663	
664	
665	
666	

667 Acknowledgements

668 This work utilized the computational resources of the high-performance computation Biowulf 669 cluster at National Institutes of Health, USA (http://hpc.nih.gov).

670 Female Lung Cancer Consortium in Asia (NCI): This study was supported by a Grant-in-Aid for 671 Scientific Research on Priority Areas from the Ministry of Education, Science, Sports, Culture 672 and Technology of Japan, a Grant-in- Aid for the Third Term Comprehensive 10-Year Strategy 673 for Cancer Control from the Ministry Health, Labor and Welfare of Japan, by Health and Labor 674 Sciences Research Grants for Research on Applying Health Technology from the Ministry of 675 Health, Labor and Welfare of Japan, by the National Cancer Center Research and Development 676 Fund, the National Research Foundation of Korea (NRF) grant funded by the Korea government 677 (MEST) (grant No. 2011-0016106), a grant of the National Project for Personalized Genomic 678 Medicine, Ministry for Health & Welfare, Republic of Korea (A111218-11-GM04), the Program 679 for Changjiang Scholars and Innovative Research Team in University in China (IRT 14R40 to 680 K.C.), the National Science & Technology Pillar Program (2011BAI09B00), MOE 111 Project 681 (B13016), the National Natural Science Foundation of China (No. 30772531, and 81272618), 682 Guangdong Provincial Key Laboratory of Lung Cancer Translational Medicine (No. 683 2012A061400006), Special Fund for Research in the Public Interest from the National Health 684 and Family Planning Commission of PRC (No. 201402031), and the Ministry of Science and 685 Technology, Taiwan (MOST 103-2325-B-400-023 & 104-2325-B-400-012). The Japan Lung 686 Cancer Study (JLCS) was supported in part by the Practical Research for Innovative Cancer 687 Control from Japan Agency for Medical Research and Development (15ck0106096h0002) and 688 the Management Expenses Grants from the Government to the National Cancer Center (26-A-1) 689 for Biobank. BioBank Japan was supported by the Ministry of Education, Culture, Sports, 690 Sciences and Technology of the Japanese government. The Japan Public Health Center-based 691 prospective Study (the JPHC Study) was supported by the National Cancer Center Research and 692 Development Fund (23-A- 31[toku], 26-A-2, 29-A-4, 2020-J-4, and 2023-J-4) (since 2011) and a 693 Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan 694 (from 1989 to 2010). The Taiwan GELAC Study (Genetic Epidemiological Study for Lung 695 AdenoCarcinoma) was sup- ported by grants from the National Research Program on Genomic 696 Medicine in Taiwan (DOH99-TD-G-111-028), the National Research Program for 697 Biopharmaceuticals in Taiwan (MOHW 103-TDUPB-211-144003, MOST 103-2325-B-400-023)

698 and the Bioinformatics Core Facility for Translational Medicine and Biotechnology 699 Development (MOST 104-2319-B-400-002). This work was also supported by the Jinan Science 700 Research Project Foundation (201102051), the National Key Scientific and Technological 701 Project (2011ZX09307-001-04), the National Natural Science Foundation of China 702 (No.81272293), the State Key Program of National Natural Science of China (81230067), the 703 National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) 704 (No. NRF- 2014R1A2A2A05003665), Sookmyung Women's University Research Grants, Korea 705 (1-1603-2048), Agency for Science, Technology and Research (A*STAR), Singapore and the 706 US National Institute of Health Grant (1U19CA148127-01). The overall GWAS project was 707 supported by the intramural program of the US National Institutes of Health/National Cancer 708 Institute. The following is a list of grants by study center: SKLCS (Y.T.K.)—National Research 709 Foundation of Korea (NRF) grant funded by the Korea government (MEST) (2011-0016106). 710 (J.C.) – This work was supported by a grant from the National R&D Program for Cancer 711 Control, Ministry of Health & Welfare, Republic of Korea (grant no. 0720550-2). (J.S.S) – grant 712 number is A010250. WLCS (T.W.)—National Key Basic Research and Development Program 713 (2011CB503800). SLCS (B.Z.)—National Nature Science Foundation of China (81102194). 714 Liaoning Provincial Department of Education (LS2010168). China Medical Board (00726). GDS 715 (Y.L.W.)—Foundation of Guangdong Science and Technology Department (2006B60101010, 716 2007A032000002, 2011A030400010). Guangzhou Science and Information Technology Bureau 717 (2011Y2-00014). Chinese Lung Cancer Research Foundation, National Natural Science 718 Foundation of China (81101549). Natural Science Foundation of Guangdong Province 719 (S2011010000792). TLCS (K.C., B.Q)—Program for Changjiang Scholars and Innovative 720 Research Team in University (PCSIRT), China (IRT1076). Tianjin Cancer Institute and Hospital. 721 National Foundation for Cancer Research (US). FLCS (J.C.W., D.R., L.J.)-Ministry of Health 722 (201002007). Ministry of Science and Technology (2011BAI09B00). National S&T Major 723 Special Project (2011ZX09102-010-01). China National High-Tech Research and Development 724 Program (2012AA02A517, 2012AA02A518). National Science Foundation of China 725 (30890034). National Basic Research Program (2012CB944600). Scientific and Technological 726 Support Plans from Jiangsu Province (BE2010715). NLCS (H.S.)—China National High-Tech 727 Research and Development Program Grant (2009AA022705). Priority Academic Program 728 Development of Jiangsu Higher Education Institution. National Key Basic Research Program

729 Grant (2011CB503805). GEL-S (A.S.)—National Medical Research Council Singapore grant 730 (NMRC/0897/2004, NMRC/1075/2006). (J.Liu)—Agency for Science, Technology and 731 Research (A*STAR) of Singapore. GELAC (C.A.H.)—National Research Program on Genomic 732 Medicine in Taiwan (DOH98-TDG-111-015). National Research Program for 733 Biopharmaceuticals in Taiwan (DOH 100- TD-PB-111-TM013). National Science Council, 734 Taiwan (NSC 100- 2319-B-400-001). YLCS (Q.L.)—Supported by the intramural pro- gram of 735 U.S. National Institutes of Health, National Cancer Institute. SWHS (W.Z., W-H.C., N.R.)-The 736 work was supported by a grant from the National Institutes of Health (R37 CA70867, UM1 737 CA182910) and the National Cancer Institute intramural research program, including NCI 738 Intramural Research Program contract (N02 CP1101066). JLCS (K.M., T.K.)-Grants-in-Aid 739 from the Ministry of Health, Labor, and Welfare for Research on Applying Health Technology 740 and for the 3rd-term Comprehensive 10-year Strategy for Cancer Control; by the National 741 Cancer Center Research and Development Fund; by Grant-in-Aid for Scientific Research on 742 Priority Areas and on Innovative Area from the Ministry of Education, Science, Sports, Culture 743 and — Technology of Japan. (W.P.)—NCI R01-CA121210. HKS (J.W.)— General Research 744 Fund of Research Grant Council, Hong Kong (781511M). The Environment and Genetics in 745 Lung Cancer Etiology (EAGLE), Prostate, Lung, Colon, Ovary Screening Trial (PLCO), and 746 Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) studies were supported by the 747 Intramural Research Program of the National Institutes of Health, National Cancer Institute 748 (NCI), Division of Cancer Epidemiology and Genetics. ATBC was also supported by U.S. Public 749 Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-RC-37004) from the NCI. 750 PLCO was also supported by individual contracts from the NCI to the University of Colorado 751 Denver (NO1-CN-25514), Georgetown University (NO1-CN-25522), the Pacific Health 752 Research Institute (NO1-CN-25515), the Henry Ford Health System (NO1-CN-25512), the 753 University of Minnesota, (NO1-CN-25513), Washington University (NO1-CN-25516), the 754 University of Pittsburgh (NO1-CN-25511), the University of Utah (NO1-CN-25524), the 755 Marshfield Clinic Research Foundation (NO1-CN-25518), the University of Alabama at 756 Birmingham (NO1-CN-75022), Westat, Inc. (NO1-CN-25476), and the University of California, Los Angeles (NO1-CN-25404). The Carotene and Retinol Efficacy Trial (CARET) is funded by 757 758 the National Cancer Institute, National Institutes of Health through grants U01-CA063673, 759 UM1-CA167462, and U01-CA167462. The Cancer Prevention Study-II (CPS-II) Nutrition

760	Cohort was supported by the American Cancer Society. The NIH Genes, Environment and
761	Health Initiative (GEI) partly funded DNA extraction and statis- tical analyses (HG-06-033-NCI-
762	01 and RO1HL091172-01), genotyping at the Johns Hopkins University Center for Inherited
763	Disease Research. This research was supported by the National Research Foundation of Korea
764	(NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C4002236)
765	
766	
767	Female Lung Cancer Consortium in Asia (Tianjin): Tianjin Science and Technology Committee
768	Foundation, 18YFZCSY00520.
769	
770	Female Lung Cancer Consortium in Asia (Taiwan): The Ministry of Health and Welfare grants
771	DOH97-TD-G-111-028 (ISC), DOH98-TD-G-111-017 (ISC), DOH99-TD-G-111-014 (ISC);
772	DOH97-TD-G-111-026 (CAH), DOH98-TD-G-111-015 (CAH), DOH99-TD-G-111-028
773	(CAH); National Health Research Institutes grants NHRI-PH-110-GP-01, NHRI-PH-110-GP-03;
774	and the Ministry of Science and Technology grants MOST108-2314-B-400-038(CAH),
775	MOST109-2740-B-400-002(CAH), MOST 111-2740-B-400-002 (CAH), MOST 111-2314-B-
776	400-020 (CAH).
777	
778	The GWAS of lung cancer in European never smokers was supported by NIH R01 CA149462
779	(OYG).
780	
781	OncoArray study in Europeans: The OncoArray data and analysis from INTEGRAL-ILCCO
782	were supported by NIH U19 CA203654, and U19 CA148127. The data harmonization for
783	ILCCO was supported by Canadian Institute for Health Research (CIHR) Canada Research Chair
784	to R.J.H, and CIHR FDN 167273).
785	
786	European never-smoking lung cancer study: CIA is a Research Scholar of the Cancer Prevention
787	Institute of Texas (CPRIT) and supported by CPRIT grant RR170048.
788	
789	Taiwan eQTL study: This study was supported by the Ministry of Health and Welfare grants
790	DOH97-TD-G-111-028 (ISC), DOH98-TD-G-111-017 (ISC), DOH99-TD-G-111-014 (ISC);

- 791 DOH97-TD-G-111-026 (CAH), DOH98-TD-G-111-015 (CAH), DOH99-TD-G-111-028
- 792 (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),
- 794 MOST109-2740-B-400-002(CAH).
- 795 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
- 797 Research Funds, NIH-CA77118 and CA80127. G.L. is supported is supported by the Alan
- 798 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.
- 800
- 801

802 Author contributions

- 803 B.B., N.R., Q.L. and H.Z. organized and designed the study. B.B., K.W., H.Z. carried out the
- statistical analyses. B.B., K.W., Q.L. N.R. and H.Z. wrote the first draft. Q.L., M.T.L., B.A.B.,
- 805 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.,
- 806 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,
- 807 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,
- 808 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.,
- 809 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.,
- 810 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.,
- 811 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.,
- 812 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.,
- 813 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.,
- 814 YH.T., WC.W., KY.C., MS.H., WC.S., YM.C., CL.W., KC.C., CJ.Y., HH.H., FY.T., HC.L.,
- 815 CJ.C., PC.Y., K.Shiraishi, T.K., H.K., S.M., H.H., K.Goto, Y.Ohe, S.W., Y.Yatabe, M.T.,
- 816 R.Hamamoto, A.Takahashi, Y.Momozawa, M.Kubo, Y.K., Y.D., Y.Miyagi, H.N., T.Y., N.S.,
- 817 M.I., M.H., Y.N., K.Takeuchi, K.W., K.Matsuda, Y.Murakami, K.S., K.T., Y.O., M.S.,
- 818 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.,
- 819 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.,
- 820 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.,
- 821 A.G.S., R.H., M.R.S., O.Y.G., I.P.G., X.W., P.Y. conducted epidemiology studies and

- 822 contributed samples to GWAS and/or conducted initial genotyping. All authors reviewed and
- 823 approved the final review of the manuscript.

825 Competing interests

- 826 The authors declare no competing interests.

- 0.57

842 Table 1. Prediction performance of different methods for generating polygenic risk scores

843 for lung cancer in never-smoking East Asian populations.

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

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval; AUC, area

- 845 under the curve; PC, principal component; CT, clumping and thresholding
- ¹ORs are provided per one unit standard deviation of PRS.
- ²Adjusted for age (continuous) and first 10 principal components.
- ³Sample size for single-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian
- 849 (EAS) ancestry for training, 2,219 cases and 2,272 controls of EAS ancestry for tuning and 2,219 cases
- and 2,272 controls of EAS ancestry for validation.
- ⁴SNPs that reached genome-wide significance in Shi et al., 2023.
- ⁵SNPs obtained from genome-wide association study conducted in individuals with European ancestry
- 853 (Hung et al., 2021, Cancer Res.).
- ⁶Sample size for multi-ancestry PRS methods included 3,564 cases and 16,238 controls of East Asian
- (EAS) ancestry, as well as 2,058 cases and 5,575 controls of European ancestry for training, 2,219 cases
- and 2,272 controls of EAS ancestry for tuning and 2,219 cases and 2,272 controls of EAS ancestry for
- 857 validation.
- 858

	Genetic variance of PRS ²	Proportion of all-GWAS	Proportion of Familial
		variants genetic variance	risk explained by
Sample size ¹		explained by PRS ³	PRS ⁴
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%

859 Table 2. Genetic variance in East Asian lung cancer among never smokers explained by

860 LDpred2 PRS

861 ¹ Sample Size: The current sample comprises 3,564 cases and 16,238 controls. Future projections assume

a 1:10 case-control ratio for sample sizes of 35,000 and 55,000 cases. Genetic variance projections for the

LDpred2 PRS are based on the GENESIS method, originally designed for the CT method (refer to Nat.

64 Genet. 50, 1318-1326 (2018)), extended to include LDpred2 by modeling the variance ratio between

865 LDpred2 and CT (Supplementary Note).

866 ² *Genetic Variance of PRS*: This corresponds to the heritability on the frailty scale, assuming a polygenic

867 log-additive model underpins this relationship. It quantifies the proportion of the phenotype variation that868 can be attributed to genetic factors in the context of PRS.

³ Proportion of Genetic Variance from All-GWAS Variants Explained by PRS: This represents the

870 variance of all genome-wide imputable variants as established through LD-score regression (refer to Nat.

871 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 $\sigma_{GWAS}^2 = Var(\sum_{m=1}^{M} \beta_m G_m)$, where G_m is the standardized genotype

873 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

- 875 variants is 0.755.
- ⁴ Proportion of Familial Risk Explained by PRS: This calculates the familial risk in terms of genetic
- variance using the formula $\lambda_s^2 = \exp(\sigma^2)$, where λ_s is the familial risk when a first-order sibling has the
- 878 disease, and σ^2 is the genetic variance on frailty-scale. Further details of this calculation can be found in
- 879 Nat. Genet. 31, 33-36 (2002). For lung cancer in East Asian never-smokers, the familial risk is a 1.84-fold
- 880 increase, and the genetic variance of all GWAS variants, as estimated through LD-score regression,
- explains 61.9% of this increased familial relative risk.



884

Figure 1. Overview of data structure, polygenic risk score (PRS) development, validation
and application. Summary statistics from East Asian (EAS) genome-wide association studies
(GWAS) were used to develop single-ancestry PRS using methods such as a simple PRS
constructed using 25 SNPs that have previously reached genome-wide significance (i.e., P<5x10⁻
⁸) (PRS-25), a PRS using the clumping and thresholding (CT) method (PRS-CT) incorporating 8
SNPs, and a PRS using a genome-wide Bayesian-based approach, LDpred2 (LDpred2 PRS)

891	incorporating close to a million SNPs. For the multi-ancestry PRS development, we also used
892	summary statistics from European (EUR) GWAS, applying the PRS-CSx method that leveraged
893	genome-wide association summary statistics for close to a million SNPs with a Bayesian
894	continuous shrinkage prior to model SNP effect sizes across populations, as well as CT-SLEB
895	method, which enhances the standard CT methods with a two-dimensional approach to select
896	SNPS for EAS PRS construction by incorporating over 2 million SNPs. Tuning and validation of
897	each PRS was conducted in an independent EAS individual-level data. Relative risk per PRS
898	quantile was calculated as an odds ratio (OR) with the middle quantile (40 th to 60 th percentile) set
899	as the reference, and the area under the receiver operating curve (AUC) was estimated for each
900	PRS. CT-SLEB PRS was used to estimate 10-year and lifetime cumulative absolute risk, and
901	PRS-CT and PRS-LDpred2 were used for sample size projection.
902	
903	
904	
905	
906	
907	
908	
909	
910	
911	
912	
913	
914	
915	
916	
917	
918	
919	





926



931 PRS-25 (A), (B) Clumping and thresholding method, PRS-CT, (C) Bayesian-based genome-wide

- 932 approach, LDpred2 PRS, and (D) multi-ancestry approach, CT-SLEB.
- 933

934 A



937

939 Figure 3. Association between polygenic risk score (PRS) and lung adenocarcinoma by age

- 940 groups. (A) Odds ratios (ORs) per standard deviation (SD) of the PRS and 95% confidence
- 941 intervals, and (B) ORs for individuals in the upper 90th percentile of the PRS.



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 package³⁸, based on Taiwan's age-specific incidence and mortality data, and the PRS relative risks, as described in the Methods section.



Figure 5. Lifetime cumulative absolute risk of developing lung adenocarcinoma by family history. Lifetime (age 30-80) absolute risk of developing lung adenocarcinoma in never-smoking East Asian women stratified by first-degree family history of lung cancer. Absolute risks were calculated using iCARE package³⁸, based on Taiwan's age specific incidence and mortality data, and the relative risks of the PRS and family history, as described in the Methods section.





Figure 6. 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.

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 neversmokers. *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 largescale 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 casecontrol 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. Hu, Z. *et al.* A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.* **43**, 792–796 (2011).
- 29. Wang, L. *et al.* Genetically determined height was associated with lung cancer risk in East Asian population. *Cancer Med.* **7**, 3445–3452 (2018).
- 30. Dong, J. *et al.* Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat. Genet.* **44**, 895–898 (2012).
- 31. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- 32. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
- 33. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).
- 34. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 35. Hung, R. J. *et al.* Assessing Lung Cancer Absolute Risk Trajectory based on a Polygenic Risk Model. *Cancer Res* **81**, 1607 (2021).
- 36. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- 37. Foucher, Y. *et al.* RISCA: Causal Inference and Prediction in Cohort-Based Analyses. *R* package version 1.01 Preprint at https://cran.r-project.org/package=RISCA (2020).
- 38. Pal Choudhury, P. *et al.* iCARE: An R package to build, validate and apply absolute risk models. *PLoS One* **15**, e0228198 (2020).
- 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).
- 40. Zhang, Y., Qi, G., Park, J. H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).

- 41. Choudhury, P. P. *et al.* iCARE: An R package to build, validate and apply absolute risk models. *PLoS One* **15**, (2020).
- 42. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
- 43. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, (2019).
- 44. Kullo, I. J. *et al.* Polygenic scores in biomedical research. *Nat. Rev. Genet.* **23**, 524–532 (2022).
- 45. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
- 46. 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).
- 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).
- 48. Schaid, D. J., Sinnwell, J. P., Batzler, A. & McDonnell, S. K. Polygenic risk for prostate cancer: Decreasing relative risk with age but little impact on absolute risk. *The American Journal of Human Genetics* **109**, 900–908 (2022).
- 49. 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).
- 50. 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).
- 51. 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).
- 52. Chen, T. *et al.* Genomic Insights for Personalized Care: Motivating At-Risk Individuals Toward Evidence-Based Health Practices. *medRxiv* 2024.03.19.24304556 (2024) doi:10.1101/2024.03.19.24304556.

Data Availability

All data supporting the findings described in this paper are available in the paper and in the Supplementary Information and from the corresponding author or as otherwise indicated upon request. The individual genotype data for the FLCCA data are in dbGaP phs000716.v1.p1 (Genome-Wide Association Study of Lung Cancer Susceptibility in Never-Smoking Women in Asia). For the NCC and ACC studies, please contact Kouya Shiraishi at kshirais@ncc.go.jp or Takashi Kohno at tkkohno@ncc.go.jp for summary statistics. The GWAS data for the European populations contributing to this study are available at dbGap under accession phs000877.v1.p1

(Transdisciplinary Research Into Cancer of the Lung (TRICL), <u>https://www.ncbi.nlm.nih.gov/projects/gap/cgibin/study.cgi?study_id=phs000876.v2.p1</u>), phs001273.v3.p2 (Oncoarray Consortium, <u>https://www.ncbi.nlm.nih.gov/projects/gap/cgibin/study.cgi?study_id=phs001273.v3.p2</u>). To gain access to all data in dbGaP cited in this paper, please apply for dbGaP Authorized Access. GWAS data for the European populations contributing to this study are available under accession phs000877.v1.p1 (ILLCO), phs001273.v1.p1 (OncoArray).

Code Availability

Methods implemented and their corresponding repositories include: SCT and LDpred2 at <u>https://github.com/privefl/bigsnpr</u>, PRS-CSx at <u>https://github.com/getian107/PRScsx</u>. CT-SLEB at <u>https://github.com/andrewhaoyu/CTSLEB</u>, and GENESIS at <u>https://github.com/yandorazhang/GENESIS</u>. PLINK: <u>https://www.cog-genomics.org/plink/1.9</u>. Most of our statistical analyses were performed using the following R packages: ggplot2 v.3.3.3, dplyr v.1.0.4, data.table v.1.13.6, iCARE v.1.30.0

https://github.com/KevinWFred/PRS_EASLC