medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Large-Scale Gene-Smoking Interactions and Fine Mapping Study Identifies Multiple Novel Blood Pressure Loci in over 1 Million Individuals

Mengyu Zhang1*, Michael R. Brown2*, Amy R. Bentley3*, Thomas W. Winkler4*, Raymond Noordam5,133*, Pavithra Nagarajan6*, Bohong Guo1, Songmi Lee7, Karen Schwander8, Wenyi Wang9, Kenneth Westerman10,134, Jeffrey R. O'Connell11, Farah Ammous12, Anna D. Argoty-Pantoja13, Traci M. Bartz14, Chiara Batini15,135, Palwende R. Boua16,99,178, Heather J. Cordell17, Laia Díez-Ahijado18, Latchezar Dimitrov19, Anh N. Do20, Jiawen Du21, Mary F. Feitosa8, Ayush Giri22,136, Franco Giulianini23, Valborg Gudmundsdottir24,137, Xiuqing Guo25, Sarah E. Harris26, Natalie R. Hasbani27, Janina M. Herold4, Keiko Hikino28, Edith Hofer29,138, Fang-Chi Hsu30, Anne U. Jackson31, Minjung Kho32, Aldi T. Kraja33, Leo-Pekka Lyytikäinen34,139, Aline Meirhaeghe35, Manon Muntaner35, Masahiro Nakatochi36, Giuseppe Giovanni Nardone37, Teresa Nutile38, Nicholette D. Palmer19, Alessandro Pecori39, Varun S. Rao40, Rainer Rauramaa41, Mihir M. Sanghvi42,140, Aurora Santin37,39, Botong Shen43, Heather M. Stringham31, Fumihiko Takeuchi44, Ye An Tan45, Jingxian Tang46, Sébastien Thériault47, Olga Trofimova48, Stella Trompet49, Peter J. van der Most13, Ya Xing Wang50, Zhe Wang51,141, Yujie Wang52, Erin B. Ware12, Stefan Weiss53,142, Ananda Rajitha Wickremasinghe54, Chenglong Yu55, Wanying Zhu56, Md Abu Yusuf Ansari57, Pramod Anugu58, Bernhard Banas59, R. Graham Barr60, Til B. Basnet61,114, Eric Boerwinkle2,143, Carsten A. Böger59,144, Max Breyer62, Ulrich Broeckel63, Luke Bryant64,135, Brian Cade6,145, Silvia Camarda37, Pasqualina Cennamo38, William Checkley65, Miao-Li Chee66, Guanjie Chen3, Kayesha Coley15, Stacey D. Collins12, Martin H. de Borst67, Lisa de las Fuentes20,146, Ian J. Deary26, Charles T. Dupont68, Christian Enzinger69,147, Tariq Faquih6,145, Jessica D. Faul12, Lilian Fernandes Silva70, Victoria Gauthier35, Adam D. Gepner71, Mathias Gorski4, Hans Jörgen Grabe72,148, Mariaelisa Graff52, Charles C. Gu20, Jiang He73, Sami Heikkinen74, Bertha Hidalgo75, Heather M. Highland52, James E. Hixson27, Michelle M. Hood76, Steven C. Hunt77, Marguerite Ryan Irvin75, Masato Isono78, Mika Kähönen79,149, Sharon L. R Kardia76, Carrie A. Karvonen-Gutierrez76, Anuradhani Kasturiratne54, Tomohiro Katsuya80, Joel D. Kaufman81, Heikki A. Koistinen82,150,179, Pirjo Komulainen41, Bernhard K. Krämer83,151, Lenore J. Launer84, Hampton Leonard85, Daniel Levy86,152, Lifelines Cohort Study87, Jianjun Liu88, Pedro Marques-Vidal89, Angel Martinez-Perez18,153, Koichi Matsuda90, John J. McNeil55, Yuri Milaneschi91, J. Jaime Miranda92,154, John L. Morrison93, Michael Nalls85, Maggie CY Ng94, Ilja M. Nolte13, Jill M. Norris95, Anniina Oravilahti70, Amit Patki96, Lauren E. Petty56, Patricia A. Peyser76, Giulia Pianigiani37,39, Laura M. Raffield97, Olli T. Raitakari98,155, Michele Ramsay99, Ken M. Rice100, Paul M. Ridker23,156, Lorenz Risch101,157, Martin Risch102,158, Daniela Ruggiero38, Edward A. Ruiz-Narvaez103, Tom C. Russ26,159,180, Charumathi Sabanayagam66, Nataraja Sarma Vaitinadin22, Reinhold Schmidt29, Laura J. Scott31, Surina Singh99,160, Colleen M. Sitlani104, Roelof A.J Smit51,161, Jennifer A. Smith12,76, Quan Sun21, E Shyong Tai45,105, Kent D. Taylor25, Paola Tesolin37,39, Yih Chung Tham66,162, The Biobank Japan Project90, Chikowore Tinashe106, Rob M. van Dam45,107, Julien Vaucher89,163, Uwe Völker53,142, Chaolong Wang108, Otis D. Wilson61,164, Tien-Yin Wong66,165, Jianzhao Xu19, Ken Yamamoto109, Jie Yao25, Mitsuhiro Yokota110, Kristin L. Young52, Martina E. Zimmermann4, Philippe Amouyel35, Jennifer E. Below56, Sven Bergmann48, Antonio Bernabe-Ortiz92, Laura Bierut111, Michael Boehnke31, Donald W. Bowden19, Jean-Tristan Brandenburg99,166, Daniel I. Chasman23,156, Ching-Yu Cheng66,162, Marina Ciullo38, Maria Pina Concas39, David Conen112, Simon R. Cox26, Luc Dauchet35, Hithanadura Janaka de Silva54, Marcus Dörr113,142, Todd L. Edwards114,136, Ervin R. Fox58, Nora Franceschini115,167, Barry I. Freedman116, Giorgia Girotto37,39, Vilmundur Gudnason24,137, Sioban D. Harlow76, Iris M. Heid117, Adriana M. Hung61,164, Sahoko Ichihara118, Cashell E. Jaquish119, Catherine John15,135,181, Jost B. Jonas120,168, J. Wouter Jukema121,169, Norihiro Kato78, Bernard D. Keavney122,170, Tanika N. Kelly123, Markku Laakso70,171, Paul Lacaze55, Timo Lakka41,74,182, Seunggeun Lee32, Terho Lehtimäki34,139, Ching-Ti Liu46, Ruth J.F Loos51,161, Kari E. North52, Brenda Penninx91, Michael A. Province8, Bruce M. Psaty104,172,183, Susan Redline6,145, Frits R. Rosendaal5, Charles N. Rotimi3, Jerome I. Rotter25, Maria Sabater-Lleal18,153,184, Helena Schmidt124, Xueling Sim45, Harold Snieder13, Beatrice Spedicati37,39, Klaus

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

J. Stark4, Chikashi Terao125,173,185, Lynne E. Wagenknecht126, David R. Weir12, Wei Zhao12,76, Xiaofeng Zhu127, Patricia B. Munroe128, Yan V. Sun129,174, James Gauderman93, Alisa K. Manning130,175, Myriam Fornage7, Hugues Aschard131,176, Heming Wang6,145#, Paul S. de Vries2#, Gao Wang132,177#, Dabeeru C. Rao20#, Alanna C. Morrison2#, Han Chen2#

*These authors contributed equally
#These authors jointly directed this work

Author Affiliations:
1. Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA
2. Human Genetics Center, Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA
3. Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD, USA
4. Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany
5. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
6. Division of Sleep and Circadian Disorders, Department of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA
7. McGovern Medical School, Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA
8. Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA
9. Department of Human Genetics, Leiden University Medical Center, Leiden, South Holland, the Netherlands
10. Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, USA
11. Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA
12. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA
13. Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
14. Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, Seattle, WA, USA
15. Genetic Epidemiology Group, Department of Population Health Sciences, University of Leicester, Leicester, UK
16. Clinical Research Unit of Nanoro, Institut de Recherche en Sciences de la Santé, Ouagadougou, Kadiogo, Burkina Faso
17. Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
18. Unit of Genomics of Complex Disease, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Catalonia, Spain
19. Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC, USA
20. Center for Biostatistics and Data Science, Institute for Informatics, Data Science, and Biostatistics, Washington University in St. Louis, St. Louis, MO, USA
21. Department of Biostatistics, University of North Carolina at Chapel Hill, NC
22. Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA
23. Division of Preventive Medicine, Medicine, Brigham and Women's Hospital, Boston, MA, USA
24. Icelandic Heart Association, Kopavogur, Iceland
25. Genomic Outcomes, Pediatrics, The Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA
26. Lothian Birth Cohorts, Psychology, The University of Edinburgh, Edinburgh, UK
27. Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

28. Laboratory for Pharmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan
29. Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria
30. Division of Public Health Sciences, Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC, USA
31. Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
32. Graduate School of Data Science, Seoul National University, South Korea
33. University of Mississippi Medical Center, Jackson, MS, USA
34. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
35. Lille University; Lille University Hospital; INSERM U1167 - RID-AGE; Institut Pasteur de Lille, University of Lille, Inserm, Institut Pasteur de Lille, Lille, France
36. Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan
37. Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy
38. Institute of Genetics and Biophysics "A. Buzzati-Traverso", CNR, Naples, Italy
39. Institute for Maternal and Child Health, I.R.C.C.S. “Burlo Garofolo”, Trieste, Italy
40. Graduate Program in Life Sciences, University of Maryland, Baltimore, Baltimore, MD, USA
41. Kuopio Research Institute of Exercise Medicine, Kuopio, Finland
42. William Harvey Research Institute, Centre for Pharmacology and Precision Medicine, Queen Mary University of London, London, UK
43. National Institute on Aging, Bethesda, MD, USA
44. Department of Bioinformatics, National Institute of Global Health and Medicine, Japan Institute for Health Security, Shinjuku-ku, Tokyo, Japan
45. Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore, Singapore
46. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
47. Department of Molecular Biology, Medical Biochemistry and Pathology, Université Laval, Quebec City, Quebec, Canada
48. Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
49. Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands
50. Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Sciences Key Laboratory, Beijing, China
51. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
52. Gillings School of Global Public Health, Department of Epidemiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
53. Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
54. Faculty of Medicine, University of Kelaniya, Kelaniya, Ragama, Sri Lanka
55. School of Public Health and Preventive Medicine, Monash University, Victoria 3004 Australia
56. Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
57. Department of Data Science, University of Mississippi Medical Center, Jackson, MS, USA
58. Jackson Heart Study, University of Mississippi Medical Center, Jackson, MS, USA
59. Department of Nephrology, University Hospital Regensburg, Regensburg, Germany
60. General Medicine, Medicine, Columbia University Medical Center, New York, NY, USA
61. Tennessee Valley Healthcare System, Nashville VA Medical Center, Nashville, TN, USA
62. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
63. Section of Genomic Pediatrics, Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA
64. Department of Respiratory Sciences, University of Leicester, Leicester, UK
65. Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

66. Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore, Singapore
67. Nephrology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
68. Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA
69. Division of General Neurology, Department of Neurology, Medical University of Graz, Graz, Austria
70. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland
71. Cardiovascular Medicine, Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA
72. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany
73. Department of Epidemiology, O'Donnell School of Public Health, University of Texas Southwestern Medical Center, Dallas, TX, USA
74. Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus, Kuopio, Finland
75. Epidemiology, UAB School of Public Health, Birmingham, Alabama, USA
76. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA
77. Epidemiology, Internal Medicine, University of Utah, Salt Lake City, UT, USA
78. Department of Gene Diagnostics and Therapeutics, National Institute of Global Health and Medicine, Japan Institute for Health Security, Shinjuku-ku, Tokyo, Japan
79. Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland
80. Department of Clinical Gene Therapy, Graduate School of Medicine, Osaka University, Suita-shi, Osaka, Japan
81. Epidemiology, University of Washington Seattle, Seattle, WA, USA
82. Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland
83. Fifth Department of Medicine (Nephrology/Endocrinology/Rheumatology/Pneumology), University Medical Centre Mannheim, University of Heidelberg, Mannheim, Germany
84. Intramural Research Program, Laboratory of Epidemiology and Population Sciences, Intramural Research Program, Laboratory of Epidemiology and Population Sciences
85. NIH Center for Alzheimer's and Related Dementias, Bethesda, MD, USA
86. Framingham Heart Study, Framingham, MA, USA
87. Lifelines Biobank, Roden, the Netherlands
88. Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore, Singapore
89. Internal Medicine, Department of Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
90. Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan
91. Psychiatry, Amsterdam, the Netherlands
92. CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru
93. Division of Biostatistics, Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, USA
94. Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
95. Department of Epidemiology, Colorado School of Public Health, Aurora, CO, USA
96. Biostatistics, UAB School of Public Health, Birmingham, Alabama, USA
97. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA
98. Centre for Population Health Research, Turku University Hospital, Turku, Finland
99. Sydney Brenner Institute for Molecular Bioscience (SBIMB), University of the Witwatersrand, Johannesburg, Gauteng, South Africa
100. Department of Biostatistics, University of Washington, Seattle, WA, USA
101. Institute of Laboratory Medicine, Faculty of Medical Sciences, Private University of the Principality of Liechtenstein, Triesen, Liechtenstein
102. Dr. Risch Medical Laboratory, Vaduz, Liechtenstein
103. Department of Nutritional Sciences, University of Michigan, Ann Arbor, MI, USA
104. Cardiovascular Health Research Unit, General Internal Medicine, Department of Medicine, University of Washington, Seattle, WA, USA

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

105. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore, Singapore
106. Channing Division of Network Medicine, Department of Medicine, Brigham Hospital, Harvard University, Boston, Massachusetts, USA
107. Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA
108. School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hebei, China
109. Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, Fukuoka, Japan
110. Kurume University School of Medicine, Kurume, Fukuoka, Japan
111. Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA
112. Population Health Research Institute, McMaster University, Hamilton, ON, Canada
113. Cardiology, Pneumology, Infectious Diseases, Intensive Care Medicine, Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany
114. Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
115. Department of Epidemiology, University of North Carolina at Chapel Hill, NC
116. Section on Nephrology, Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA
117. national Institute of Global Health and Medicine, University of Regensburg, Regensburg, Germany
118. Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, Tochigi, Japan
119. Division of Cardiovascular Science, Epidemiology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA
120. Privatpraxis Prof Jonas und Dr Panda-Jonas, Heidelberg, Germany, Germany
121. Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands
122. Cardiovascular Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK
123. Division of Nephrology, Department of Medicine, University of Illinois Chicago, Chicago, IL, USA
124. Division of Genetic Epidemiology, Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria
125. Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan
126. Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA
127. Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA
128. Clinical Pharmacology and Precision Medicine, Queen Mary University of London, London, UK
129. Rollins School of Public Health, Department of Epidemiology, Emory University, Atlanta, GA, USA
130. Clinical and Translational Epidemiology Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
131. Department of Computational Biology, Institut Pasteur, Université Paris Cité, Paris, France
132. Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA
133. Health Campus the Hague/Public Health and Primary Care, Leiden University Medical Center, the Hague, the Netherlands
134. Department of Medicine, Harvard Medical School, Boston, MA, USA
135. University Hospitals of Leicester NHS Trust, Leicester, UK
136. VA Tennessee Valley Healthcare System (626), Nashville, TN, USA
137. Faculty of Medicine, University of Iceland, Reykjavik, Iceland
138. Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria
139. Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Department of Clinical Chemistry, Tampere University, Tampere, Finland
140. Barts Heart Centre, Barts Health NHS Trust, London, UK

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

141. Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA
142. DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany
143. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
144. Department of Nephrology, Hospital Traunstein, Traunstein, Germany
145. Division of Sleep and Circadian Disorders, Department of Medicine, Harvard Medical School, Boston, MA, USA
146. Cardiovascular Division, Department of Medicine, Washington University in St. Louis, St. Louis, MO, USA
147. Division of Neuroradiology, Interventional and Vascular Radiology, Department of Radiology, Medical University of Graz, Graz, Austria
148. German Centre for Neurodegenerative Diseases, Site Rostock/Greifswald, Greifswald, Germany
149. Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Department of Clinical Physiology, Tampere University, Tampere, Finland
150. University of Helsinki, Research Programs Unit, Clinical and Molecular Metabolism; Department of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
151. European Center for Angioscience ECAS, Medical Faculty Mannheim of the University of Heidelberg, Mannheim, Germany
152. Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA
153. Centre for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain
154. Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
155. Research Centre of Applied and Preventive Cardiovascular Medicine, Department of Clinical Physiology and Nuclear Medicine, University of Turku, Turku, Finland
156. Harvard Medical School, Boston, MA
157. Center of Laboratory Medicine, University Institute of Clinical Chemistry, University of Bern, Bern, Switzerland
158. Zentrallabor, Kantonsspital Graubünden, Chur, Switzerland
159. Alzheimer Scotland Dementia Research Centre, Psychology, The University of Edinburgh, Edinburgh, UK
160. Division of Human Genetics, School of Pathology, University of the Witwatersrand, Johannesburg, Gauteng, South Africa
161. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
162. Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore, Singapore
163. Division of Internal Medicine, Department of Medicine and Specialties, Fribourg Hospital and University of Fribourg, Fribourg, Switzerland
164. Division of Hypertension and Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
165. Tsinghua Medicine, Tsinghua University, Haidian District, Beijing, China
166. Strengthening Oncology Services Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng, South Africa
167. Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, NC
168. Rothschild Foundation Hospital, Paris, France
169. Netherlands Heart Institute, Utrecht, the Netherlands
170. Manchester Heart Institute, Manchester University NHS Foundation Trust, Manchester, UK
171. Department of Medicine, Kuopio University Hospital, Kuopio, Finland
172. Department of Epidemiology, University of Washington, Seattle, WA, USA
173. Clinical Research Center, Shizuoka General Hospital, Shizuoka, Shizuoka, Japan
174. Veterans Affairs Atlanta Healthcare System, Decatur, GA, USA
175. Metabolism Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA
176. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
177. Department of Neurology, Columbia University, New York, NY, USA
178. MRC Unit The Gambia at London School of Hygiene and Tropical Medicine, Banjul, The Gambia

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

179. Minerva Foundation Institute for Medical Research, Helsinki, Finland
180. Division of Psychiatry, Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK
181. Leicester British Heart Foundation Centre of Research Excellence, Leicester, UK
182. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland
183. Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA
184. Cardiology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden
185. Department of Applied Genetics, School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Shizuoka, Japan

Corresponding Author:
Han Chen, PhD
Human Genetics Center, Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030, USA
Email: Han.Chen.2@uth.tmc.edu
Tel: +1 (713) 500-9958

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Abstract

Cigarette smoking influences blood pressure (BP) levels. Studying and accounting for potential gene-smoking interactions can help discover novel loci and provide insights into biological pathways for smoking-associated BP regulation. We conducted a genome-wide association meta-analysis involving 1,188,241 individuals from 66 studies in five ancestry groups, analyzing systolic BP, diastolic BP, and pulse pressure while considering interactions between genetic variants and three smoking exposures: smoking status, cigarettes per day, and pack years. These analyses identified twelve novel loci for BP at genome-wide significance ($P < 5 \times 10^{-9}$), and highlighted biological processes including tight junction integrity, mitochondrial health, vascular relaxation, and endothelial function. In smoking status-stratified analyses, smoking modifies the genetic effect of six variants on BP. To prioritize likely causal, we developed and applied SuSiEgxe, a fine-mapping method based on a two-degree-of-freedom joint test using gene-environment interaction summary statistics. Fine-mapped loci uncovered immune-related pathway for smoking-associated BP regulation.

Keywords: GWAS, blood pressure, gene by smoking interaction, joint test, fine mapping

1 Introduction

The exploration of genetic variants related to blood pressure (BP) regulation has been significantly advanced through genome-wide association studies (GWAS). Gene-by-environment (G×E) interaction studies are critical for understanding how genetic predispositions may amplify or mitigate the effects of lifestyle choices. Among these lifestyle exposures, cigarette smoking is a major risk factor for high BP (hypertension) and has been shown to influence BP regulation through various biological mechanisms. Given that hypertension is a core risk factor for cardiovascular diseases, smoking may contribute to cardiovascular risk directly or through its impact on BP[1, 2, 3]. Investigating gene-by-smoking interactions in BP GWAS can reveal novel loci and enhance understanding of the biological pathways that regulate BP. Particularly, it is crucial to identify BP variants with heterogeneous associations in smokers and nonsmokers or across different levels of smoking exposure, e.g., cigarettes per day or pack years. Such analyses reveal genetic factors that may influence BP differently according to smoking status and helps to refine the understanding of how environmental exposures like smoking interact with genetic variation.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Previous studies have investigated gene-smoking interactions in BP traits, particularly within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. However, they did not include marginal genetic effect-only analyses that excluded smoking, and it was unknown if any of the loci found through the two-degree-of-freedom (2df) joint test could have been discovered without smoking or gene-smoking interaction in the model[4]. This gap makes it difficult to determine whether the loci newly identified through the 2df joint test could have been detected without accounting for smoking or gene-smoking interactions. Moreover, there was no analysis with fine mapping to determine the causal single-nucleotide polymorphisms (SNPs) underlying the observed interactions. Fine mapping aims to narrow down the potential causal variants by analyzing linkage disequilibrium (LD) patterns (CAVIAR[5], SuSiE[6, 7], FINEMAP[8], etc.) and functional annotations (eCAVIAR[9], PAINTOR[10], etc.). However, existing fine mapping methods were developed focusing on the marginal genetic effect only. Fine-mapping approaches within the context of gene-environment interactions remain underdeveloped, especially using genomic summary statistics.

In this study, we conducted genome-wide association meta-analyses incorporating gene-smoking interactions to identify genetic loci associated with BP traits and to assess the modifying effect of cigarette smoking on the genetic architecture of BP. We also conducted fine mapping with our developed tool SuSiEgxe that uses summary statistics from the 2df joint test in variant-smoking status (CURSMK) interaction meta-analyses. Our analysis included a large sample size (1,188,241 individuals from five ancestry groups) to maximize statistical power for discovery. To account for potential heterogeneity in interaction effects, we report results stratified by ancestry group, in addition to our primary cross-population meta-analyses.

2 Results

The study received approval from the Institutional Review Board of The University of Texas Health Science Center at Houston and was conducted in compliance with all applicable ethical regulations. Ethical approval for data collection was obtained from the relevant review boards of each participating cohort, and informed consent was obtained from all participants.

2.1 Overview

The investigation includes meta-analyses of 66 studies, and 1,188,241 samples from five ancestry groups: 804,275 (68%) European ancestry (EUR), 246,825 (21%) East Asian ancestry

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

(EAS), 86,580 (7%) African ancestry (AFR), 37,949 (3%) Hispanic admixed ancestry (HIS), 12,612 (1%) South Asian ancestry (SAS). Sex-stratified meta-analyses were performed to evaluate sex heterogeneity in the associations as well. We focused on systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse pressure (PP) while considering interactions between genetic variants and three smoking exposures: current smoking status (CURSMK), and within current smokers, cigarettes per day (CPD), and pack years (PY). Meta-analyses of all combinations of smoking exposures and BP traits were conducted within each ancestry and across all five ancestries (Cross-Population Meta-Analysis; CPMA). The descriptive statistics of the studies can be found in Supplementary Table S1. We considered two models for each smoking-BP combination within each population group separately. Model 1 is a joint analysis of the genetic main effect and smoking interaction effect where we considered a one-degree-of-freedom (1df) interaction test for the interaction term and a 2df joint test that considered the main and interaction term. Model 2 only analyzed the marginal genetic effect without accounting for smoking exposures.

With the criteria described in the Methods section, we identified 1,753 independent loci by accounting for interaction effects with CURSMK, CPD or PY from either 1df interaction test or 2df joint test in ancestry specific and cross-population analyses. No loci were identified in SAS. The 1df interaction tests identified two novel BP loci and ten significant known BP loci (Table 1-2). The 2df joint tests identified ten novel BP loci (Table 1) and 106 known BP loci (Supplementary Table S6) that were not identified by marginal tests in Model 2 (see details in Methods section). Among the 12 novel loci in total that reached a stringent genome-wide significance threshold ($P < 5 \times 10^{-9}$), two were for SBP, six for DBP, and four for PP (Supplementary Figure S1).

Variant-smoking interaction and joint analyses in CPMA

In the combined sex analyses (Supplementary Figure S2 A), the 1df interaction test revealed four loci showing variant-smoking interaction effects on BP outcomes, including two novel loci: rs35975869 (*AC021021.2*; $P_{GxE} = 4.1 \times 10^{-10}$) for the variant-CPD interaction effect on PP, and rs61840542 (*FRMPD2*; $P_{GxE} = 4.7 \times 10^{-9}$) for the variant-CURSMK interaction effect on SBP. There was no evidence of a difference in either of these interactions by sex. In the 2df joint test, 826 loci were identified, and nine were novel (Table 1). In males, the 2df joint test revealed two novel BP loci: rs72832924 (*GRB7*; $P_{2df} = 2.8 \times 10^{-10}$) showed the joint effect of variant-CURSMK interaction on DBP and was only identified in males, and rs143429637 (*RP11-813P10.2*; $P_{2df} = 2.1 \times 10^{-24}$) showed the joint effect of variant-CPD interaction on PP. There

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

were no associations for 1df interaction among males or females, and no novel BP loci were identified with the 2df joint test among females (Supplementary Table S4).

Variant-smoking interaction and joint analyses in EUR

In the EUR combined sex analyses (Supplementary Figure S2 B), one novel locus with lead variant rs35975869 (*AC021021.2*; $P_{GxE} = 1.5 \times 10^{-9}$) was identified by the 1df interaction test in the analysis of variant-CPD interaction on PP. In the 2df joint test, rs143429637 (*RP11-813P10.2*; $P_{2df} = 1.1 \times 10^{-16}$) was a novel BP locus revealed by the analysis of variant-CPD interaction on PP. These loci were also identified in CPMA. In males, 292 statistically significant loci were identified in the 2df joint test, and one of them with lead variant rs143429637 (*RP11-813P10.2*; $P_{2df} = 3.2 \times 10^{-27}$) was novel. There were no associations for the 1df interaction among males, and no novel BP loci were identified with the 2df joint test among females.

Variant-smoking interaction and joint analyses in EAS

There were no novel associations from the 1df interaction or 2df joint test in EAS. In the combined sex analyses (Table 2, Supplementary Figure S2 C), the known BP locus with lead variant rs3782886 (*BRAP*) was identified in variant-CURSMK interaction for both SBP ($P_{GxE} = 7.0 \times 10^{-24}$) and PP ($P_{GxE} = 1.5 \times 10^{-11}$). The known BP locus with lead variant rs116873087 (*NAA25*) was identified in variant-CURSMK interaction on DBP ($P_{GxE} = 7.0 \times 10^{-19}$).

Variant-smoking interaction and joint analyses in AFR

There were no novel associations from the 1df interaction or 2df joint test in AFR. The 2df joint test revealed nine significant known BP loci in the combined sex analyses (Supplementary Figure S2 D). Two known BP loci with lead variants rs1716694 (*ULK4*; $P_{2df} = 6.4 \times 10^{-10}$) and rs74884236 (*MFSD1*; $P_{2df} = 4.4 \times 10^{-9}$) were only found in females (Supplementary Table S7).

Variant-smoking interaction and joint analyses in HIS

There were no novel associations from the 1df interaction or 2df joint test in HIS. The 2df joint test found two known BP loci in the combined sex analyses (Supplementary Figure S2 E). The variant rs935168 (*KCNK3*; $P_{2df} = 2.1 \times 10^{-12}$) was identified in variant-CURSMK interaction study on DBP, and rs2193950 ($P_{2df} = 2.2 \times 10^{-9}$) was identified in the variant-CURSMK interaction with SBP (Supplementary Table S7).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2.2 Smoking status-stratified analyses for significant 1df interaction variants

We further investigated the smoking-status stratified associations for the significant 1df interaction variants with the framework developed by Laville, et al[11] and applied with tool **j2s** (https://gitlab.pasteur.fr/statistical-genetics/j2s). The results (Table 3) demonstrated differences in the genetic effects on blood pressure between current smokers and non-current smokers. rs3782886-T is a missense variant at *BRAP*, and it has a significant association with higher PP among smokers ($\beta = 1.09$, $P = 6.9 \times 10^{-22}$), but not among non-smokers ($\beta = 0.27$, $P = 3.9 \times 10^{-8}$) in EAS. Similarly, rs116873087-C and rs12231737-T are intronic variants at *NAA25*, and *TRAFD1* respectively. The two loci have significant association with lower DBP among smokers but not among non-smokers in females and males. rs11066280-A is an intronic variant at *HECTD4*, and it has a significant association with lower PP among smokers but not among non-smokers. Additionally, rs4237919-A and rs1327235-A are at a gene for long intergenic non-protein coding RNA and are significantly associated with PP and SBP among non-smokers, respectively.

2.3 Functional annotations for novel loci

The gene mappings for *LGSN* (rs11293017), *TSHZ2* (rs6123158), *RPL39L* (rs111563399), *GSDMB* (rs72832924), and *ORMDL3* (locus 12) were validated by tissue-specific eQTL evidence from GTExv8, which was observed in whole blood, lung, or subcutaneous adipose tissue (Supplementary Table S12). Tissue-specific (GTExv8) eQTL associations were observed at rs61840542 (Table 1, locus 1) in artery tibial tissues. Additionally, the expression of *STARD* and *GRB7* were associated with the novel locus with lead variant rs72832924 (Table 1, locus 12), which was exclusively identified in males, based on eQTL evidence from GTExv8 cells cultured fibroblasts tissue. Similarly, the gene mappings for *PTCH1* (9:95507136:CG_C), *TSHZ2* (rs6123158), and *GLB1* (rs34408570) were observed in cells cultured fibroblasts tissue based on eQTL GTExv8 evidence.

Many genes were linked to novel loci through chromatin interactions (Supplementary Table S12). These interactions were observed in various tissues and cell types, including the left and right ventricles of the heart, aorta, lung, GM12878 (lymphoblastoid cell line), IMR90 (lung fibroblast cell line), mesenchymal stem cells, and mesendoderm cells.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

We performed gene-set enrichment analysis using genes mapped through positional, expression quantitative trait loci (eQTL), and chromatin interaction data (Supplementary Figure S3). The carbohydrate phosphatase activity gene set was associated with the novel locus with lead variant 9:95507136:CG_C ($\text{adjusted } P = 0.036$). The 17q12 Copy Number Variation Syndrome gene set was linked to the novel locus with lead variant rs8070260 (*MYO19*; $\text{adjusted } P = 7.8 \times 10^{-69}$). Additionally, GWAS catalog gene sets related to asthma were mapped to the novel locus with lead variant rs72832924, which was identified specifically in males.

2.4 New findings from 2df fine mapping

Causal variants were identified by our new 2df fine mapping tool SuSiEgxe (Supplementary Table S13), followed by the gene mapping and functional annotations by FUMA. We investigated one novel locus on gene *MYO19*, and five loci from previously reported genes associated with BP (Supplementary Table S3): *KCNK3, ACAD10, MAEA, CCDC162P, and ASB3* (Supplementary Table S14).

For the known BP locus with lead variant rs12476527 at gene *KCNK3,* which was also reported in our previous work[4], fine mapping narrowed the region (2:26681997-26710019) from 28 kb to 18.9 kb and identified a 13-variant credible set from 27 significant variants in EAS (Figure 1A), and narrowed the region (2:26673079-26714801) from 41.7 kb to 11.6 kb and identified a nine-variant credible set from 30 significant variants in CPMA (Figure 1B). Causal variants were mapped to genes *KCNK3, NRBP1*, and *CENPA* by eQTL evidence in heart atrial appendage tissue and heart left ventricle tissue. Chromatin interactions were observed in GM12878 (lymphoblastoid cell line), IMR90 (lung fibroblast cell line), mesenchymal stem cells, and mesendoderm cells (Supplementary Table S14). The gene-set enrichment analysis highlighted the pathways of fatty acid β-oxidation, mitochondrial metabolism, and valproic acid metabolism (Supplementary Figure S4 A).

The novel *MYO19* locus with lead variant rs8070260 was identified by the 2df joint test in Model 1 ($P_{2df} = 1.4 \times 10^{-9}$), whereas it was not identified through the marginal test in Model 2 ($P_{1df} = 8.3 \times 10^{-9}$). Fine mapping analysis using the 2df joint test pinpointed three potential causal variants, including rs1109442, rs8070260, and rs7222903 (Supplementary Figure S5 A), which

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

exhibited consistent functional annotations when incorporating all variants within the locus (Supplementary Table S14, Supplementary Figure S4 B).

Similarly, loci with lead variants rs60722337 at *MAEA*, rs6902892 at *CCDC162P*, and rs74598641 at *GPR75-ASB3 / ASB3* (Supplementary Figure S5 B-D) were identified by the 2df joint test in Model 1 but were not detected by the marginal test in Model 2. Causal variants were mapped to genes *MAEA, UVSSA, CRIPAK, CTBP1, CD164, SMPD2, AK9, FIG4, PSME4, GPR75*, and *GPR75-ASB3* by eQTL evidence in tissue whole blood, artery aorta, artery coronary, artery tibial, heart atrial appendage, heart left ventricle, lung, and cells cultured fibroblasts. Chromatin interactions were observed in various tissues and cell types, including the left and right ventricles of the heart, aorta, lung, GM12878 (Lymphoblastoid cell line), IMR90 (lung fibroblast cell line), mesenchymal stem cells, and mesendoderm cells (Supplementary Table S14).

The locus with lead variant rs11066015 at gene *ACAD10* was only identified in the East Asian population (EAS) (Supplementary Figure S5 E). Chromatin interactions between causal variants and genes *ATXN2, BRAP, ACAD10, ALDH2, MAPKAPK5,* and *TRAFD1* were observed in IMR90 (lung fibroblast cell line), mesenchymal stem cells, and mesendoderm cells (Supplementary Table S14). The mapped genes show a highly significant enrichment for traits linked to coronary artery disease, coronary heart disease, hypertension, mean arterial pressure, and alcohol consumption (Supplementary Figure S4 C).

2.5 Genome-wide MAGMA analysis and gene-set enrichment

MAGMA gene-based and gene-set tests were applied to the full list of 2df joint p-values of CPMA for all smoking and BP combinations in the combined sex analyses (Supplementary Table S5).

We observed 19 significant protein-coding genes mapped based on 2df joint p-values from multiple analyses (Supplementary Table S8). The gene *ULK4* was mapped in five analyses: DBP-CURSMK ($P = 1.2 \times 10^{-17}$), DBP-CPD ($P = 1.8 \times 10^{-6}$), DBP-PY ($P = 2.3 \times 10^{-6}$), PP-CPD ($P = 4.8 \times 10^{-7}$), and PP-PY ($P = 2.8 \times 10^{-8}$). Additionally, the genes *HECTD4*, *NAA25*, and *TRAFD1* were found based on the tests of variant-smoking interaction with traits SBP and DBP.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

MAGMA gene-set analyses (Supplementary Table S9) highlighted several gene sets and biological processes that affect the association between smoking and BP. Significant gene sets were related to the regulation of RNA metabolic processes (DBP-CURSMK, $P = 1.1 \times 10^{-15}$), calcium channel activity in cardiac muscle cells (DBP-CURSMK, $P = 4.5 \times 10^{-15}$; SBP-CURSMK, $P = 4.0 \times 10^{-7}$), muscle structure development (PP-CURSMK, $P = 5.0 \times 10^{-14}$), muscle tissue development (PP-CURSMK, $P = 9.9 \times 10^{-13}$), and elastic fiber assembly (PP-CURSMK, $P = 3.1 \times 10^{-12}$).

Gene set enrichment analyses using Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA)[12] GENE2FUNC (Supplementary Table S10) highlighted several biological processes that affect the blood, heart, or lungs. Selected significant pathways were the regulation of BP (DBP-CURSMK, adjusted $P = 0.015$; SBP-CURSMK, adjusted $P = 0.019$), tube morphogenesis (PP-CURSMK, adjusted $P = 0.011$), and regulation of tube size (DBP-CURSMK, adjusted $P = 0.004$; SBP-CURSMK, adjusted $P = 0.033$), heart development, regulation of heart contraction, heart valve morphogenesis, the circulatory system processes and development with genes *ACE, AGT, NOS3, NPPA, NPPB*, etc. involved. Notably, the pathways related to type I interferon in the immune system were found in SBP, DBP, and quantitative smoking exposure CPD and PY analyses with genes *PTPN11, OAS1, OAS3,* and *OAS2*.

We assessed whether the positionally mapped genes were overrepresented in traits and diseases previously identified in the GWAS Catalog based on a 2df joint test in CPMA with the GENE2FUNC module in FUMA (Supplementary Table S11). Traits associated with SBP, DBP, PP, hypertension, and mean arterial pressure demonstrate highly significant enrichment across various smoking exposure groups with genes such as *KCNK3, ULK4, BRAP*, etc. Additionally, coronary artery disease and myocardial infarction exhibit significant enrichment across multiple gene sets with genes such as *CYP17A1, CNNM2, ACAD10, ALDH2, NAA25, HECTD4*, etc.

3 Discussion

This study presents a comprehensive analysis of gene-smoking interactions on BP across five ancestry groups. A total of 66 studies and 1,188,241 participants were included in the analysis.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Including EUR, EAS, AFR, HIS, and SAS allowed for a broad exploration of genetic associations. Our study identified 12 novel loci, associated with BP traits, with a focus on SBP (n = 2 loci), DBP (n = 6 loci), and PP (n = 4 loci).

Newly identified protein-coding genes showed potential pathways for regulating BP through the interaction with smoking. The GxE 1df interaction test identified a novel locus (Table 1) adjacent to *FRMPD2* (rs61840542, $P_{GxE} = 4.7 \times 10^{-9}$), which is a scaffolding protein that maintains epithelial cell polarity and tight junction integrity to support correct cell-cell adhesion and barrier function[13] and in glutamate-gated ion channels expressed in the central nervous system[14]. The disruption of *FRMPD2*-mediated signaling may lead to increased permeability in the lungs and blood vessels, and then promotes inflammation and disease progression through impaired tight junctions and epithelial integrity triggered by smoking[15].

The GxE 1df interaction test identified 10 BP loci (Table 2) that were reported to have marginal effects in previous studies and highlighted their effect on BP with interaction of smoking. The locus with lead variant rs56654546 at *PPT1* was identified in EUR females. The locus with lead variant rs78572043 at *CUX2* was identified in EAS males. Loci with lead variants rs3782886 at *BRAP*[16], rs116873087 at *NAA25*[17, 18], and rs12231737 at *TRAFD1*[18] were only identified in EAS. A locus with lead variant rs11066280 at *HECTD4*[19, 20] was identified in CPMA. These loci are located near the protein-coding genes known to be related to BP, including *COL9A2*[21], *PPTC7*[18], *MYL2*[19], *SH2B3*[22], *ATXN2*[17, 23], *ACAD10*[17], *ALDH2*[20], *ERP29*[24], *RPL6-ALDH1*[20], *PDE3A*[25], *SLX4IP*[25], *JAG1*[25, 26, 27], *BANF2*[25], *MAL2*[24], *SNORA32*[28], *MIR548AZ - CCN3*[17], *NOV*[26], *CCN3 - ENPP2*[17], *ENPP2*[18]. *CUX2* was associated with DBP and SBP through N^6-methyladenosine (m^6A)-associated SNP rs7398833 in East Asian population[29]. In a study of elderly Japanese population that accounted for smoking status, the minor allele of rs3782886 at *BRAP* was inversely associated with hypertension only in participants with high hematopoietic activity[30]. *NAA25,* in a region with *ERP29*[31], was previously reported to be associated with SBP and DBP[32].

The smoking status-stratified analyses further illustrate how smoking modifies the genetic effect on BP. One significant 1df variant rs3782886-T at *BRAP* is associated with increased BP among smokers, while three loci at *NAA25*, *TRAFD1* and *HECTD4* were associated with decreased BP among smokers. *TRAFD1* is a protein-coding gene that is primarily related to immune and inflammatory responses via its interactions with NF-κB and TNF receptor signaling

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

pathways[33, 34]. Cigarette smoking activates the NF-kB pathway, which up-regulates various pro-inflammatory genes, leading to increased levels of inflammatory markers and cytokines, and contributing to its pro-inflammatory effects on the immune system[35], potentially increasing the susceptibility to hypertension and cardiovascular diseases through inflammatory mechanisms and vascular dysfunction. *HECTD4* is a protein-coding gene involved in protein breakdown and recycling. It helps regulate the process of tagging proteins for degradation through the ubiquitin system. Shencun Fang, et al[36]. identified a potential role of the circHECTD1/HECTD1 pathway in regulating endothelial-mesenchymal transition (EndMT), where both endothelial and epithelial cells contribute to fibroblast accumulation, ultimately promoting the development of pulmonary fibrosis in silicosis, and cardiovascular diseases.

The GxE 2df joint model newly identified protein-coding genes *MYO19* (rs8070260, $P_{2df} = 1.4 \times 10^{-9}$), *VRK2* (rs1106090, $P_{2df} = 2.8 \times 10^{-9}$), *NDUFC1* (rs140470470, $P_{2df} = 1.2 \times 10^{-10}$), *RPL39L* (rs111563399, $P_{2df} = 1.3 \times 10^{-9}$), and *GRB7* (rs72832924, $P_{2df} = 2.8 \times 10^{-10}$ only in males). They showed potential pathways for regulating BP through the joint effect of genetic main effect and interaction with smoking (Table 1). A novel locus with the lead variant rs13032423 mapped to *VRK2*, was previously reported to be associated with BP through an interaction with long total sleep time[37]. Along with the novel locus with lead variant rs1106090 discovered in this study, these findings highlight *VRK2* as potentially relevant to cardiovascular health. *GRB7* has been previously associated with asthma, another trait influenced by smoking exposure[38, 39]. *MYO19* functions as a protein that controls both the movement and operations of mitochondria inside cells. Maintaining mitochondrial health is crucial for essential energy production and cellular stress response mechanisms. Smoking can cause oxidative stress alongside inflammation with the potential to harm mitochondrial structures. Since *MYO19* plays a role in mitochondrial dynamics, disruptions in its function could worsen the effects of smoking and hypertension on cellular health, particularly in the cardiovascular system. This pathway has been verified in mouse experiments by Sergey Dikalov, et al[40]. *NDUFC1* is a subunit of mitochondrial complex I, which plays a key role in the electron transport chain and ATP production. ATP is necessary for maintaining various cellular functions, including the regulation of endothelial nitric oxide synthase (eNOS), which is responsible for nitric oxide (NO) synthesis, a key player in vascular relaxation and endothelial function[41]. Dysfunction in *NDUFC1* may reduce the efficiency of the electron transport chain, leading to decreased ATP production, which can exacerbate endothelial dysfunction and

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

smooth muscle cell impairment in blood vessels, contributing to the progression of high BP. *RPL39L* also influences cell proliferation by regulating protein synthesis and mitochondrial function[42]. Furthermore, we have discovered one molecular function involving the novel locus with lead variant 9:95507136:CG_C that is related to carbohydrate phosphatase activity (Supplementary Figure S3 A). Enzyme activities involve catalyzing the removal of phosphate groups from carbohydrates or carbohydrate derivatives. When mitochondrial structures are damaged due to smoking-induced oxidative stress, it may disrupt the normal activity of these enzymes, which results in altered metabolic processes and contributes to the pathogenesis of smoking-related diseases, including metabolic disorders, and eventually leads to elevation in BP.

There were 106 loci identified through GxE 2df joint model that were not significant in the marginal genetic tests without accounting for smoking, which supports they were driven by gene by smoking interactions (Supplementary Table S6). Some loci are at genes *PRDM16*[17, 43] (rs111063488), *NOS3*[28, 43, 44] (rs117564322), *CREB3L1*[45] (11:46323702:G_GA), *GSDME*[17] (rs2521769), *NPR1*[28, 43] (rs35479618), *PRKD3*[28, 46] (rs59916529), and *YEATS2*[47] (rs262969). *PRDM16* deficiency is associated with an increased risk of cardiomyopathy and cardiac mortality[48], and its deficiency in vascular smooth muscle cells leads to disrupted BP circadian variation[49]. *NOS3* (*E298D* and *T-786C* polymorphisms) was previously reported to be associated with the risk of incident coronary heart disease with the interaction of cigarette smoking[50]. Recently, a study showed a positive impact of *NOS3* on angiogenesis through enhanced endothelial cells derived from induced pluripotent stem cells (iPSC-ECs)[51]. Additionally, loss of *CREB3L1* expression leads to upregulation of *FGFbp1* and pleiotrophin, both of which positively regulate angiogenesis[52]. Therefore, *NOS3* and *CREB3L1* are potentially associated with hypertension through the pathway of angiogenesis regulation[53, 54]. *GSDME*-mediated pyroptosis is associated with inflammation of lung[55] and atherosclerosis[56], and there is a strong link between atherosclerosis and hypertension[57]. *NPR1* in vascular endothelial cells is the receptor for natriuretic peptides and thus an important regulator of BP via salt/water balance[58]; *NPR1* deficiency also leads to augmented expression of pro-inflammatory cytokines[59].

Several genes were discovered through the MAGMA and gene mapping and have been associated with BP (Supplementary Table S8-S11). *ULK4*, one of the genes found in five analyses for DBP and PP and other GWAS, encodes a member of the unc-51-like

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

serine/threonine kinase (STK) family and plays a role in neuronal growth and endocytosis. It was reported in many GWAS for BP and hypertension[22, 60, 61, 62]. *HECTD4* and *TRAFD1* were also found in multiple MAGMA gene-based analyses. We also found several genes from novel loci tied to fibroblasts based on eQTL GTEx/v8 evidence, such as *PTCH1*, *TSHZ2*, *GLB1*, *STARD3*, and *GRB7* (Supplementary Table S12).

In this study, we developed and applied fine mapping to the 2df joint test, allowing us to pinpoint the set of potential causal SNPs, particularly for loci that were not identified in the 1df marginal tests at a significance threshold of $P < 5 \times 10^{-9}$. For instance, the novel locus with lead variant rs8070260 in *MYO19* was identified through the joint analysis of the genetic main effect and the variant-CURSMK interaction effect ($P_{2df} = 1.4 \times 10^{-9}$) but was not detected in the marginal genetic effect analysis ($P_{1df} = 8.3 \times 10^{-9}$). This underscores the value of fine mapping based on the 2df joint test, which leverages its higher statistical power after accounting for heterogeneous genetic effects by smoking exposures. Our 2df fine mapping narrowed down 116 (89.9%) association regions out of 129 selected loci (Supplementary Table S13). More importantly, our 2df fine mapping method showed multi-ancestry fine mapping resulted in narrower credible sets, likely due to the increased stringency of LD when combining data across populations. Out of 29 loci on which we performed both CPMA and single ancestry fine mapping, 19 (65.5%) loci showed a narrower credible set in the multi-ancestry fine mapping.

Based on the causal variants in 95% credible sets, we found more potential pathways or related genes in the regulation of BP and the interaction with smoking. A novel locus with lead variant rs8070260 in *MYO19* was mapped to genes based on genomic position and chromatin interaction, and the gene-set enrichment analysis indicates the mapped genes were related to 17q12 copy number variation (CNV) syndrome (Supplementary Figure S4 A). 17q12 CNV syndrome has been linked to diabetes and kidney disease, both of which are related to hypertension and cardiovascular disease[63, 64, 65]. The genetic deletions or duplications in this region may affect pathways involved in vascular function, potentially increasing susceptibility to high BP and heart disease.

We investigated three more loci that were only discovered in the 2df joint test and not significant in the marginal genetic association test (M2). They were located at rs60722337 (*MAEA*), rs6902892 (*CCDC162P*), and rs74598641 (*GPR75-ASB3* / *ASB3*). The causal variants were mapped to genes *RNF212*, *GPR75*, etc. *MAEA* gene functions in transcriptional regulation while

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

impacting the cell cycle and apoptosis[66, 67] and may play a role in cancer development and immune system functioning. *CCDC162P* was reported to be associated with human red blood cell traits[68]. *ASB3* is involved in immune regulation, cell signaling, and the degradation of specific proteins, mainly through its interaction with the ubiquitin-proteasome system[69, 70].

Our study benefits from a large sample size and multiple population groups, allowing us to identify both population-specific and shared genetic effects, and the LD structure in multi-ancestry populations provides narrower credible sets in fine mapping compared to population-specific fine mapping. Our study utilizes advanced statistical methods, including 2df fine mapping, to improve the resolution of genetic signals. Additionally, functional annotation and enrichment analyses provide mechanistic insights into the biological pathways underlying genetic influences on BP traits with smoking interactions.

However, several limitations must be acknowledged. First, despite the large sample size, the statistical power to detect genetic effects remains limited in ancestry groups with a smaller number of participants such as South Asians and Hispanic admixed populations. Second, while we aimed to control for potential confounders such as population stratification, residual confounding may still exist, particularly for complex traits influenced by both genetic and environmental factors. Third, the quality of fine mapping in meta-analysis highly depends on the quality of LD in the reference panel used. If LD information is sourced from a reference panel that does not match the population group, this could lead to misleading results on the causal variants.

In conclusion, the identified novel loci and functional insights from enrichment analyses and fine mapping highlight multiple biological pathways involved in BP regulation, modulated by the effects of smoking. These pathways include those associated with cardiovascular complications, pulmonary function through fibroblast cells, energy metabolism (specifically mitochondrial and fatty acid metabolism), the immune system, and the broader circulatory system. Our findings underscore the significant role of genetic variation in BP regulation, particularly in the context of smoking interaction.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

4 Methods

4.1 Participating studies

Participants over 18 years old from all studies were stratified into five broad ancestry groups: European ancestry (EUR), African ancestry (AFR), Hispanic admixed (HIS), East Asian ancestry (EAS) and South Asian ancestry (SAS), and analyzed by sex groups (male, female, and combined). All studies followed a centralized analysis protocol for investigating the genome-wide gene-smoking interaction effect of BP. There were five types of studies: cross-sectional studies of unrelated subjects, cross-sectional family studies of related subjects, longitudinal studies of unrelated subjects, longitudinal family studies of related subjects, and case-control studies. All contributing studies received approval from their respective institutional review boards, and all participants provided written informed consent under the Declaration of Helsinki. Further details about the participating cohorts are provided in the Supplementary Note.

4.2 Phenotypes and Lifestyle Variables

Resting/sitting SBP (mmHg), DBP (mmHg) and PP are considered as BP outcomes. Averages of multiple readings of SBP and DBP were used in the analysis. SBP and DBP were further adjusted for subjects taking any anti-hypertensive medications by adding 15 mmHg to SBP and 10 mmHg to DBP[71]. PP values were then derived by subtracting DBP from SBP. Extreme values, more than 6 standard deviations (SD) above or below the mean, were winsorized respectively for the three BP outcomes. We considered three smoking variables, CURSMK, CPD and PY. CURSMK is a dichotomous variable and coded as 1 if a participant is a current cigarette smoker and as 0 if a participant is not a current cigarette smoker. CPD was only derived in current smokers. PY was calculated using the following formula: PY = (CPD / 20) * years of smoking. Current smokers were excluded from CPD and PY analyses if they had a value of 0 or missing for that variable. Both CPD and PY were winsorized separately in males and females after exclusion. If the number of current smokers in any group (for instance, a population group or sex stratum) was less than 100 after exclusions and winsorization, only the CURSMK analysis was conducted in that group.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

4.3 Genotype Data

Genotypes were imputed using reference panels from the Trans-Omics for Precision Medicine (TOPMed) Imputation Server, Haplotype Reference Consortium (HRC) or 1000 Genomes Project (1000G). SNPs with a minor allele frequency (MAF) less than 0.001 in the total sample (sex-combined group), low imputation quality ($r^2 < 0.3$ if using MACH or $INFO < 0.3$ if using IMPUTE2) and any SNPs mapping to sex chromosomes or mitochondria were excluded (Supplementary Table S2).

4.4 Study-specific GWAS

We considered two models for each smoking-BP combination within each population-sex group separately. Model 1 is a joint analysis of the main and interaction effects.

$$Model\ 1: E(Y) = \beta_0 + \beta_E E + \beta_G SNP + \beta_{GE} E \times SNP + \beta_C C$$

where Y is the BP phenotype; *E* is one of the smoking variables CURSMK, CPD, or PY; SNP is the dosage of the genetic variant; and *C* is a set of covariates including age, age^2, sex when analyzing males and females combined, field center (as appropriate), genetic principal components (PCs), study-specific confounders, and their interaction with smoking variables *E*, including age x E, age^2 x *E*, and sex x *E* when analyzing males and females combined. Under the null hypothesis of $\beta_{GE} = 0$, we assessed the interaction effect between the smoking variable E and genetic variants with a 1df test. Under the null hypothesis of $\beta_G = \beta_{GE} = 0$, a 2df joint test was conducted to assess the joint effect of genetic main effect and interaction effect. Model 2 only analyzes the marginal genetic effect under the null hypothesis of $\beta_G = 0$, and the same covariates are included.

$$Model\ 2: E(Y) = \beta_0 + \beta_G SNP + \beta_C C$$

Subjects with no GWAS data or missing data for any of the common covariates —including age, sex, and PCs—were excluded from the analysis.

For cross-sectional studies with unrelated subjects, model-based p-values from linear model were reported. Tools such as **LinGxEScanR** v1.0 (https://github.com/USCbiostats/LinGxEScanR), **MMAP** (https://github.com/MMAP/MMAP.github.io), **GEM v1.4.3** (https://github.com/large-scale-gxe-methods/GEM), and **EPACTS** (https://github.com/statgen/EPACTS) were utilized. A two-step approach was implemented for cross-sectional studies with related subjects: 1) linear mixed

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

models using a kinship matrix to account for family relationships were generated using **MMAP**; and 2) residuals from the linear mixed model were subsequently analyzed as the phenotype using **GEM v1.4.3**. For longitudinal studies, a single visit for each population group was chosen if it has contemporaneous measures of BP, use of anti-hypertensive medications, and smoking, to maximize the sample size and methods used as stated above. For case-control studies, all analyses were performed within case and control samples separately.

4.5 Quality Control

EasyQC2[72] (www.genepi-regensburg.de/easyqc2) was applied to perform quality control (QC) on study-level and meta-level summary statistics within each population group, sex group, and exposure type. In study-specific QC, hg19 build genomic coordinates were lifted over to hg38 genomic coordinates. Variants were excluded if they were monomorphic, had an INFO score (a measure of imputation quality) below 0.6, had an effect allele frequency outside the range of 0.01 to 0.99 with an INFO score below 0.8, or if the minor allele count for either the unexposed group (E=0) or the exposed group (E=1) was below 10. To implement a t-distribution based approach, the effective sample sizes of independent observations in the exposed group or the unexposed group N_E estimated in each study and the MAF of each genetic variant were used to compute approximate degrees of freedom (DF)[73].

4.6 Meta-analysis

For the 1df interaction test in Model 1 and marginal genetic effect in Model 2, inverse-variance weighted fixed-effects meta-analysis was performed with METAL[74] software. For the 2df joint test in Model 1, inverse-covariance-matrix weighted joint fixed-effects meta-analysis of the genetic main effect and the variant by environment interaction was performed in METAL following Manning et al[75]. All meta-analyses were conducted separately for males, females, and the combined group. They were initially performed within each population group (EUR, AFR, SAS, EAS, HIS) and subsequently across populations, with genomic control correction applied to both the population-specific meta-analyses and CPMA. The analyses excluded variants within 1 Mb of the major histocompatibility complex (MHC) region. Additionally, we removed variants with only one population contributed in CPMA and considered variants with at least 20,000 subjects in all analyses. Additional criteria were applied in population-specific meta-

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

analyses, requiring $DF_{ALL} \geq 20$, $DF_{E0} \geq 10$, $DF_{E1} \geq 10$, $INFO \geq 0.5$, $N \geq 100$ for each cohort/study, where DF is degrees of freedom approximated by N_E and MAF.

4.7 Identification of genome-wide independent loci

We used the EasyStrata2[76] package in R to identify and prioritize loci from significant findings obtained through the 1df interaction test and 2df joint test.

A variant is considered significant if it satisfies the following criteria: (1) 1df interaction test $P_{GxE} \leq 5 \times 10^{-9}$ and $FDR_{GxE} < 0.05$ or 1df interaction test $P_{GxE,m_G} \leq 0.05/m_G$ and $FDR_{GxE,m_G} < 0.05$, (2) 2df joint test $P_{2df} \leq 5 \times 10^{-9}$ and $FDR_{2df} < 0.05$. FDR_{GxE} is 1df GxE false discovery rate (FDR) calculated based on all variants genome-wide. FDR_{GxE,m_G} is 1df GxE FDR calculated based on m_G variants, and m_G represents the number of variants with marginal test $P_G \leq 10^{-5}$ from Model 2. FDR_{GxE,m_G} is 1df GxE FDR calculated based on m_G variants. FDR_{2df} is 2df joint FDR calculated based on all variants genome-wide. Significant variants and nearby variants within 500 kb regions were grouped together, and independent loci were identified by LD threshold $r^2 < 0.1$ using TOPMed-imputed 1000G reference panels. Loci were classified as novel or known by assessing their overlap with regions within 1 Mb of variants reported in previous GWAS studies for BP (see Supplementary Table S3 for known BP loci).

Heterogeneity across sexes was assessed using two-sample Z-tests under the null hypothesis of homogeneity across sexes. For each top interaction locus, two-sample Z-tests[77] were conducted to evaluate sex-specific effect size differences. A significance threshold of $P = 1 \times 10^{-6}$ was applied to these heterogeneity tests.

4.8 Smoking status-stratified analyses

For the variants identified as significant in the 1df interaction test according to the criteria described above, we conducted smoking status-stratified analyses using the tool j2s[11] (https://gitlab.pasteur.fr/statistical-genetics/j2s). Summary statistics for each variant were generated separately in smokers and non-smokers. These stratified results were used to illustrate whether smoking status modified the genetic associations with BP.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

4.9 Gene-based analyses and functional annotations

Protein-coding genes were prioritized, and additional annotations for variant mapping were obtained using FUMA v1.5.2.

MAGMA (v1.08)[78] gene-based and gene-set analyses in FUMA were conducted using the full spectrum of SNP p-values from 2df joint test. SNPs are mapped to 19,232 protein-coding genes with 10 kb window size around the genes. Genome-wide significance was defined as $P < 2.6 \times 10^{-6}$ based on Bonferroni correction. Gene sets were obtained from MsigDB v7.0, and 15,496 total gene sets were tested. Significant gene set was defined as $P < 3.23 \times 10^{-6}$ based on Bonferroni correction. To further prioritize genes, the FUMA's SNP2GENE pipeline was applied to identify candidate genes based on positional mapping. Finally, pathway enrichment analysis of the prioritized genes was performed using the FUMA's GENE2FUNC pipeline to identify enriched biological pathways, and GWAS catalog reported genes.

For novel loci, the FUMA's SNP2GENE pipeline was applied to identify candidate genes based on positional mapping (mapped to genes within 10 kb), chromatin interaction mapping ($FDR \leq 1 \times 10^{-6}$, 250 bp upstream - 500 bp downstream of transcription start site), and eQTL dataset GTEx v8. Then, gene-set enrichment analysis was performed on mapped genes using FUMA's GENE2FUNC.

4.10 Fine Mapping

We developed SuSiEgxe, a fine mapping approach that uses a 2df joint test in variant-CURSMK interaction meta-analyses using summary statistics. We applied SuSiEgxe to our meta-analysis results of SBP, DBP, and PP, accounting for genetic effect modification by smoking status (current vs. non-smoking). Fine mapping in the framework of the Sum of Single Effects (SuSiE) regression model[6, 7] was conducted in EAS, EUR, and a CPMA of five super-populations combined. Identified causal SNPs in 95% coverage credible sets were mapped to genes based on their genomic position, eQTL, and 3D chromatin interactions, followed by FUMA's GENE2FUNC gene-set enrichment analysis. More details about SuSiEgxe can be found in the Appendix.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Data Availability

The summary statistics will be made available upon publication.

Code availability

All computation central to the conclusions of this meta-analysis was conducted with the following open-source software: LinGxEScanR (https://github.com/USCbiostats/LinGxEScanR), GEM v1.4.3 (https://github.com/large-scale-gxe-methods/GEM), MMAP (https://mmap.github.io/), EPACTS (https://github.com/statgen/EPACTS), EasyQC2 and EasyStrata2 (www.genepi-regensburg.de/charge-gli), METAL (https://csg.sph.umich.edu/abecasis/metal/download/, https://genome.sph.umich.edu/wiki/Meta_Analysis_of_SNPxEnvironment_Interaction), j2s (https://gitlab.pasteur.fr/statistical-genetics/j2s), and SuSiEgxe (https://github.com/MengyuZhang1307/SuSiEgxe). Partial plots were generated by R package qqman (https://github.com/stephenturner/qqman).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Reference

1. Rahman MM, Laher I. Structural and functional alteration of blood vessels caused by cigarette smoking: an overview of molecular mechanisms. *Curr Vasc Pharmacol* **5**, 276-292 (2007).

2. Taylor JY*, et al.* A Genome-wide study of blood pressure in African Americans accounting for gene-smoking interaction. *Sci Rep* **6**, 18812 (2016).

3. Yin RX*, et al.* Interactions of several lipid-related gene polymorphisms and cigarette smoking on blood pressure levels. *Int J Biol Sci* **8**, 685-696 (2012).

4. Sung YJ*, et al.* A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. *Am J Hum Genet* **102**, 375-400 (2018).

5. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497-508 (2014).

6. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol* **82**, 1273-1300 (2020).

7. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the "Sum of Single Effects" model. *PLoS Genet* **18**, e1010299 (2022).

8. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-1501 (2016).

9. Hormozdiari F*, et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet* **99**, 1245-1260 (2016).

10. Kichaev G*, et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* **10**, e1004722 (2014).

11. Laville V*, et al.* Deriving stratified effects from joint models investigating gene-environment interactions. *BMC Bioinformatics* **21**, 251 (2020).

12. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).

13. Stenzel N, Fetzer CP, Heumann R, Erdmann KS. PDZ-domain-directed basolateral targeting of the peripheral membrane protein FRMPD2 in epithelial cells. *J Cell Sci* **122**, 3374-3384 (2009).

14. Wang G, Qi W, Liu QH, Guan W. GluN2A: A Promising Target for Developing Novel Antidepressants. *Int J Neuropsychopharmacol* **27**, (2024).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

15. Lu Q, Gottlieb E, Rounds S. Effects of cigarette smoke on pulmonary endothelial cells. *Am J Physiol Lung Cell Mol Physiol* **314**, L743-L756 (2018).

16. Yamada Y*, et al.* Identification of polymorphisms in 12q24.1, ACAD10, and BRAP as novel genetic determinants of blood pressure in Japanese by exome-wide association studies. *Oncotarget* **8**, 43068-43079 (2017).

17. Sakaue S*, et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415-1424 (2021).

18. Surendran P*, et al.* Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. *Nat Genet* **52**, 1314-1332 (2020).

19. Jeong H, Jin HS, Kim SS, Shin D. Identifying Interactions between Dietary Sodium, Potassium, Sodium-Potassium Ratios, and FGF5 rs16998073 Variants and Their Associated Risk for Hypertension in Korean Adults. *Nutrients* **12**, (2020).

20. Kato N*, et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* **43**, 531-538 (2011).

21. Keaton JM*, et al.* Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat Genet* **56**, 778-791 (2024).

22. Levy D*, et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* **41**, 677-687 (2009).

23. German CA, Sinsheimer JS, Klimentidis YC, Zhou H, Zhou JJ. Ordered multinomial regression for genetic association analysis of ordinal phenotypes at Biobank scale. *Genet Epidemiol* **44**, 248-260 (2020).

24. Zhu X, Zhu L, Wang H, Cooper RS, Chakravarti A. Genome-wide pleiotropy analysis identifies novel blood pressure variants and improves its polygenic risk scores. *Genet Epidemiol* **46**, 105-121 (2022).

25. Evangelou E*, et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet* **50**, 1412-1425 (2018).

26. Wain LV*, et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-1011 (2011).

27. Warren HR*, et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* **49**, 403-415 (2017).

28. Kichaev G*, et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

29. Mo XB, Lei SF, Zhang YH, Zhang H. Examination of the associations between m(6)A-associated single-nucleotide polymorphisms and blood pressure. *Hypertens Res* **42**, 1582-1589 (2019).

30. Shimizu Y*, et al.* Potential mechanisms underlying the association between single nucleotide polymorphism (BRAP and ALDH2) and hypertension among elderly Japanese population. *Sci Rep* **10**, 14148 (2020).

31. Brecker M, Khakhina S, Schubert TJ, Thompson Z, Rubenstein RC. The Probable, Possible, and Novel Functions of ERp29. *Front Physiol* **11**, 574339 (2020).

32. Zhang H, Mo XB, Xu T, Bu XQ, Lei SF, Zhang YH. Novel Genes Affecting Blood Pressure Detected Via Gene-Based Association Analysis. *G3 (Bethesda)* **5**, 1035-1042 (2015).

33. Sanada T, Takaesu G, Mashima R, Yoshida R, Kobayashi T, Yoshimura A. FLN29 deficiency reveals its negative regulatory role in the Toll-like receptor (TLR) and retinoic acid-inducible gene I (RIG-I)-like helicase signaling pathway. *J Biol Chem* **283**, 33858-33864 (2008).

34. van der Graaf A*, et al.* Systematic Prioritization of Candidate Genes in Disease Loci Identifies TRAFD1 as a Master Regulator of IFNgamma Signaling in Celiac Disease. *Front Genet* **11**, 562434 (2020).

35. Rom O, Avezov K, Aizenbud D, Reznick AZ. Cigarette smoking and inflammation revisited. *Respir Physiol Neurobiol* **187**, 5-10 (2013).

36. Fang S*, et al.* circHECTD1 promotes the silica-induced pulmonary endothelial-mesenchymal transition via HECTD1. *Cell Death Dis* **9**, 396 (2018).

37. Nagarajan P*, et al.* A Large-Scale Genome-Wide Study of Gene-Sleep Duration Interactions for Blood Pressure in 811,405 Individuals from Diverse Populations. *medRxiv*, (2024).

38. Herrera-Luis E*, et al.* Genome-wide association study reveals a novel locus for asthma with severe exacerbations in diverse populations. *Pediatr Allergy Immunol* **32**, 106-115 (2021).

39. Shrine N*, et al.* Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. *Lancet Respir Med* **7**, 20-34 (2019).

40. Dikalov S*, et al.* Tobacco smoking induces cardiovascular mitochondrial oxidative stress, promotes endothelial dysfunction, and enhances hypertension. *Am J Physiol Heart Circ Physiol* **316**, H639-H646 (2019).

41. Tran N, Garcia T, Aniqa M, Ali S, Ally A, Nauli SM. Endothelial Nitric Oxide Synthase (eNOS) and the Cardiovascular System: in Physiology and in Disease States. *Am J Biomed Sci Res* **15**, 153-177 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

42. Zou Q, Qi H. Deletion of ribosomal paralogs Rpl39 and Rpl39l compromises cell proliferation via protein synthesis and mitochondrial activity. *Int J Biochem Cell Biol* **139**, 106070 (2021).

43. Liu C*, et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-1170 (2016).

44. Johnson T*, et al.* Blood pressure loci identified with a gene-centric array. *Am J Hum Genet* **89**, 688-700 (2011).

45. Giri A*, et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet* **51**, 51-62 (2019).

46. Hoffmann TJ*, et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* **49**, 54-64 (2017).

47. Takeuchi F*, et al.* Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat Commun* **9**, 5052 (2018).

48. Kramer RJ*, et al.* PRDM16 Deletion Is Associated With Sex-dependent Cardiomyopathy and Cardiac Mortality: A Translational, Multi-Institutional Cohort Study. *Circ Genom Precis Med* **16**, 390-400 (2023).

49. Wang Z*, et al.* Vascular smooth muscle cell PRDM16 regulates circadian variation in blood pressure. *J Clin Invest* **135**, (2024).

50. Lee CR*, et al.* NOS3 polymorphisms, cigarette smoking, and cardiovascular disease risk: the Atherosclerosis Risk in Communities study. *Pharmacogenet Genomics* **16**, 891-899 (2006).

51. Kong AM*, et al.* NOS3 regulates angiogenic potential of human induced pluripotent stem cell-derived endothelial cells. *Biochem Biophys Rep* **40**, 101876 (2024).

52. Mellor P, Deibert L, Calvert B, Bonham K, Carlsen SA, Anderson DH. CREB3L1 is a metastasis suppressor that represses expression of genes regulating metastasis, invasion, and angiogenesis. *Mol Cell Biol* **33**, 4985-4995 (2013).

53. Humar R, Zimmerli L, Battegay E. Angiogenesis and hypertension: an update. *J Hum Hypertens* **23**, 773-782 (2009).

54. Kiefer FN, Neysari S, Humar R, Li W, Munk VC, Battegay EJ. Hypertension and angiogenesis. *Curr Pharm Des* **9**, 1733-1744 (2003).

55. Gou X*, et al.* IL-6 Prevents Lung Macrophage Death and Lung Inflammation Injury by Inhibiting GSDME- and GSDMD-Mediated Pyroptosis during Pneumococcal Pneumosepsis. *Microbiol Spectr* **10**, e0204921 (2022).

56. Qian Z*, et al.* Pyroptosis in the Initiation and Progression of Atherosclerosis. *Front Pharmacol* **12**, 652963 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

57. Chobanian AV, Alexander RW. Exacerbation of atherosclerosis by hypertension. Potential mechanisms and clinical implications. *Arch Intern Med* **156**, 1952-1956 (1996).

58. Tokudome T*, et al.* Endothelial Natriuretic Peptide Receptor 1 Play Crucial Role for Acute and Chronic Blood Pressure Regulation by Atrial Natriuretic Peptide. *Hypertension* **79**, 1409-1422 (2022).

59. Pandey KN, Vellaichamy E. Regulation of cardiac angiotensin-converting enzyme and angiotensin AT1 receptor gene expression in Npr1 gene-disrupted mice. *Clin Exp Pharmacol Physiol* **37**, e70-77 (2010).

60. Franceschini N*, et al.* Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am J Hum Genet* **93**, 545-554 (2013).

61. Natekar A, Olds RL, Lau MW, Min K, Imoto K, Slavin TP. Elevated blood pressure: Our family's fault? The genetics of essential hypertension. *World J Cardiol* **6**, 327-337 (2014).

62. Shen X, Espin-Garcia O, Qiu X, Brhane Y, Liu G, Xu W. Haplotype approach for association analysis on hypertension. *BMC Proc* **8**, S57 (2014).

63. Sanna-Cherchi S*, et al.* Copy-number disorders are a common cause of congenital kidney malformations. *Am J Hum Genet* **91**, 987-997 (2012).

64. Mefford HC*, et al.* Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet* **81**, 1057-1069 (2007).

65. Sanna-Cherchi S*, et al.* Copy-number disorders are a common cause of congenital kidney malformations. *Am J Hum Genet* **91**, 987-997 (2012).

66. Rittmann MC*, et al.* Plasma biomarkers for prediction of early tumor recurrence after resection of pancreatic ductal adenocarcinoma. *Sci Rep* **11**, 7499 (2021).

67. Wei Q, Boulais PE, Zhang D, Pinho S, Tanaka M, Frenette PS. Maea expressed by macrophages, but not erythroblasts, maintains postnatal murine bone marrow erythroblastic islands. *Blood* **133**, 1222-1232 (2019).

68. van der Harst P*, et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369-375 (2012).

69. Cheng M*, et al.* ASB3 expression aggravates inflammatory bowel disease by targeting TRAF6 protein stability and affecting the intestinal microbiota. *mBio* **15**, e0204324 (2024).

70. Hou P*, et al.* An unconventional role of an ASB family protein in NF-kappaB activation and inflammatory response during microbial infection and colitis. *Proc Natl Acad Sci U S A* **118**, (2021).

71. Newton-Cheh C*, et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* **41**, 666-676 (2009).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

72. Winkler TW*, et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-1212 (2014).

73. Sitlani CM*, et al.* Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. *Stat Med* **34**, 118-130 (2015).

74. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

75. Manning AK*, et al.* Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients. *Genet Epidemiol* **35**, 11-18 (2011).

76. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259-261 (2015).

77. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* **326**, 219 (2003).

78. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Acknowledgment

This study was largely supported by grants from the U.S. National Heart, Lung, and Blood Institute (NHLBI), the National Institutes of Health, R01HL118305, R01HL145025, and R01HL156991. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Study-specific acknowledgements are provided in the Supplementary Note.

Author Contributions

M.Z. and M.R.B. conducted centralized project data analyses, data consolidation, meta-analyses, quality control, bioinformatics analysis, and contextual interpretation. M.Z., M.R.B., A.R.B., T.W.W., R.N., P.N., H.W., P.S.de V., G.W., D.C.R., A.C.M and H.C. were part of the writing group and participated in project workflow design, interpretation of results, and drafting the manuscript. A.N.D., L.d.l.F., C.C.G., and D.C.R., participated in centralized study coordination. T.W.W., J.R.O., P.B.M., J.G., A.K.M., H.A., H.C. acted as collaborators facilitating project design, specific code scripts, or specialized analyses. All other co-authors participated in final result interpretation, cohort-level study concept and design, cohort-level phenotype data acquisition and/or quality control, cohort-level genotype data acquisition and/or quality control, and/or cohort-level data-analysis and interpretation. All authors approved the final version of the paper that was submitted to the journal.

Conflict of Interest/Disclosures:

Bruce Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Catherine John has a funded research collaboration with Orion for collaborative research projects outside the submitted work. Maria Sabater-Lleal is supported by a Miguel Servet contract from the ISCIII Spanish Health Institute (CPII22/00007) and co-

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

financed by the European Social Fund. HJG has received travel grants and speaker's honoraria from Indorsia, Neuraxpharm, Servier and Janssen. Susan Redline received personal fees from Eli Lilly and is an unpaid consultant to Apnimed. Han Chen receives consulting fees from Character Biosciences. The remaining authors declare no competing interests.

Appendix

SuSiE model with gene-environment interaction terms (SuSiEgxe)

Consider a SuSiE model with gene by environment interaction term:

$$\boldsymbol{Y} = \boldsymbol{Z\zeta} + \boldsymbol{e} = \boldsymbol{G\beta} + \boldsymbol{K\gamma} + \boldsymbol{e} \tag{1}$$

$$\boldsymbol{\zeta} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix} = \sum_{l=1}^{L} \begin{pmatrix} \boldsymbol{\beta}_l \\ \boldsymbol{\gamma}_l \end{pmatrix} \tag{2}$$

$$\boldsymbol{\zeta}_l = \sum_{l=1}^{L} \begin{pmatrix} \boldsymbol{\beta}_l \\ \boldsymbol{\gamma}_l \end{pmatrix} = \sum_{l=1}^{L} \begin{pmatrix} \boldsymbol{\xi}_l \beta_l \\ \boldsymbol{\xi}_l \gamma_l \end{pmatrix} \tag{3}$$

$$\boldsymbol{\xi}_l \sim Multinomial(1, \boldsymbol{\pi}) \tag{4}$$

$$\zeta_l = \begin{pmatrix} \beta_l \\ \gamma_l \end{pmatrix} \sim N_2(0, \boldsymbol{\Sigma}_{0l}) \tag{5}$$

$$\boldsymbol{e} = (e_1, .., e_N)^T \tag{6}$$

$$e_i \sim N(0, \sigma^2) \tag{7}$$

$\mathbf{Y}$ is an $N \times 1$ vector of phenotypes. $\boldsymbol{G} = \left(g_1, \ldots, g_{N_g}\right)$ is genotypes of N_g SNPs for n subjects, and $\boldsymbol{K} = \left(K_1, \ldots, K_{N_g}\right) = \left(g_1 E_1, \ldots, g_{N_g} E_{N_g}\right)$ is variant-environment interactions of N_g SNPs for N subjects. $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_L, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_L$ are $2L$ single-effect vectors that have only one non-zero entry (equal to β_l and γ_l respectively), and their summation is the overall effect vector $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. (β_l, γ_l)

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

follows a bivariate normal distribution with $\boldsymbol{\Sigma}_{0l}$ as the prior covariance matrix. $\boldsymbol{\xi}_l \in \{0,1\}^p$ is an indicator vector following multinomial distribution which has only one non-zero entry. $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)$ is the prior inclusion probabilities of each SNP in $\boldsymbol{G}$. e_i is an independent error terms with hyperparameter σ^2 as its variance. Assume columns of $\boldsymbol{G}$ is centered, and $\boldsymbol{G}$ and binary $\boldsymbol{E}$ are independent.

When L = 1, it becomes single-effect regression (SER) model:

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\zeta} + \boldsymbol{e} \tag{8}$$

$$\boldsymbol{\zeta} = \boldsymbol{\xi}\zeta = \begin{pmatrix} \boldsymbol{\xi}\beta \\ \boldsymbol{\xi}\gamma \end{pmatrix} \tag{9}$$

$$\boldsymbol{\xi} \sim Multinomial(1, \pi) \tag{10}$$

$$\zeta \sim N_2(0, \boldsymbol{\Sigma}_0) \tag{11}$$

$$e_i \sim N(0, \sigma^2) \tag{12}$$

IBSS-ss algorithms

The estimation of $\boldsymbol{\zeta}_l$ in equation (3) given $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{L-1}$ can be obtained by iteratively fitting a SER model with Iterative Bayesian stepwise selection using sufficient or summary statistics (IBSS-ss) (Zou et al. 2022). IBSS computes a variational approximation to the posterior distribution described below.

Posterior distribution and credible sets computation under SER model

Considering SER model equations (9-12), the posterior distribution of $\boldsymbol{\zeta} = \boldsymbol{\xi}\zeta$ is

$$\boldsymbol{\xi} \mid \boldsymbol{Z}, \boldsymbol{Y}, \sigma^2, \boldsymbol{\Sigma}_0 \sim Multinomial(1, \boldsymbol{\delta})$$

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

$$\zeta \mid \boldsymbol{Z}, \boldsymbol{Y}, \sigma^2, \boldsymbol{\Sigma}_0, \xi_j = 1 \sim N_2\left(\boldsymbol{\mu}_{1,j}, \boldsymbol{\Sigma}_{1,j}\right),\ j \in \left(1, \ldots N_g\right) \tag{13}$$

$\boldsymbol{\delta} = \left(\delta_1, \ldots, \delta_{N_g}\right)$ is the vector of posterior inclusion probability (PIPs), where δ_j $:= Pr\left(\xi_j = 1 \mid \boldsymbol{Z}, \boldsymbol{Y}, \sigma^2, \boldsymbol{\Sigma}_0\right)$. By performing N_g Bayesian simple linear regressions, the posterior mean is

$$\boldsymbol{\mu}_{1,j} = \boldsymbol{\Sigma}_{1,j} \boldsymbol{S}_j^{-1} \hat{\boldsymbol{\zeta}} \tag{14}$$

where $\boldsymbol{\Sigma}_{1,j}$ is posterior covariance, $\boldsymbol{S}_j$ is standard errors of estimates of joint effect $\hat{\boldsymbol{\zeta}}_j$ from summary data, and

$$\boldsymbol{\Sigma}_{1,j} = \left(\boldsymbol{S}_j^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1} \tag{15}$$

$$\boldsymbol{S}_j = \sigma^2 \left(\boldsymbol{Z}_j^T \boldsymbol{Z}_j\right)^{-1} \tag{16}$$

$$\hat{\boldsymbol{\zeta}}_j = \left(\boldsymbol{Z}_j^T \boldsymbol{Z}_j\right)^{-1} \boldsymbol{Z}_j^T \boldsymbol{Y} \tag{17}$$

$$\delta_j = \frac{\pi_j BF_j}{\sum_{j'=1}^{N_g} \pi_{j'} BF_{j'}} \tag{18}$$

and the Bayes Factor (BF) for comparing this model to null model ($\zeta = 0$) is

$$\begin{aligned} BF_j &:= BF\left(\boldsymbol{Z}_j, \boldsymbol{Y}; \sigma^2, \boldsymbol{\Sigma}_0\right) \\ &:= \frac{Pr\left(\boldsymbol{Y} \mid \boldsymbol{Z}_j, \sigma^2, \boldsymbol{\Sigma}_0\right)}{Pr\left(\boldsymbol{Y} \mid \boldsymbol{Z}_j; \sigma^2, \zeta = 0\right)} \\ &= \sqrt{\frac{\| \boldsymbol{\Sigma}_1 \|}{\| \boldsymbol{\Sigma}_0 \|}} exp\left(\frac{1}{2} \hat{\boldsymbol{\zeta}}_j^T \left(\boldsymbol{\Sigma}_0 + \boldsymbol{S}_j\right)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{S}_j^{-1} \hat{\boldsymbol{\zeta}}_j\right) \end{aligned} \tag{19}$$

Equations above show that the posterior distribution of $\boldsymbol{\zeta}$ can be written from $\boldsymbol{Z}^T\boldsymbol{Z}$ and $\boldsymbol{Z}^T\boldsymbol{Y}$. I denote the function that returns it as $SER - ss(\boldsymbol{Z}^T\boldsymbol{Z}, \boldsymbol{Z}^T\boldsymbol{Y}; \sigma^2, \boldsymbol{\Sigma}_0) := (\boldsymbol{\delta}, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ under SER model.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

A credible set is the set of variants with top PIPs. Variants are ranked in descending order and then added sequentially, starting from the highest PIP, until the cumulative PIP reaches the predefined 95% threshold. This set contains the most likely variants contributing to the trait with a specified probability of interest.

Sufficient statistics and approximations from GWAS meta-analysis summary data

Assume that the sample size N is large and effect sizes are small. Considering a SER model for $\xi_j = 1$, meaning only variant j has effect on Y, the score test statistics of 2df joint test is

$$\begin{aligned} T_{2df} &= \hat{\boldsymbol{\zeta}}_j^T \boldsymbol{S}_j^{-1} \hat{\boldsymbol{\zeta}}_j \\ &= N\boldsymbol{A}_j\left(1-\boldsymbol{A}_j\right)^{-1},\ \boldsymbol{A}_j = \frac{\hat{\boldsymbol{\zeta}}_j^T \boldsymbol{Z}_j^T \boldsymbol{Z}_j \hat{\boldsymbol{\zeta}}_j}{\boldsymbol{Y}^T\boldsymbol{Y}} \end{aligned} \tag{20}$$

Therefore, the residual variance is

$$\begin{aligned} \hat{\sigma}_j^2 &= \frac{\boldsymbol{Y}^T\left(1-\boldsymbol{A}_j\right)\boldsymbol{Y}}{n} \\ &= \frac{\boldsymbol{Y}^T\boldsymbol{Y}}{T_{2df}+N} \approx 1 \end{aligned} \tag{21}$$

and

$$\begin{aligned} \boldsymbol{Z}_j^T\boldsymbol{Y} &= \left(\boldsymbol{Z}_j^T\boldsymbol{Z}_j\right)\hat{\boldsymbol{\zeta}}_j \\ &= \hat{\sigma}_j^2\boldsymbol{S}_j^{-1}\hat{\boldsymbol{\zeta}}_j \end{aligned} \tag{22}$$

$\boldsymbol{Z}^T\boldsymbol{Z}$ can be approximated using LD matrix $\boldsymbol{R}$ from the reference panel. Besides the correlation between variants $\boldsymbol{g}_i$ and $\boldsymbol{g}_j$, the correlation between variants $\boldsymbol{g}_i$ and variant-environment interaction $\boldsymbol{K}_j$, correlation between $\boldsymbol{K}_i$ and $\boldsymbol{K}_j$ are considered as well. Since E is binary, $\boldsymbol{g}_i^T\boldsymbol{K}_j = \boldsymbol{K}_i^T\boldsymbol{K}_j$. From equation (16), we know

$$\begin{aligned} \boldsymbol{Z}_j^T\boldsymbol{Z}_j &= \begin{pmatrix} \boldsymbol{g}_j^T\boldsymbol{g}_j & \boldsymbol{g}_j^T\boldsymbol{K}_j \\ \boldsymbol{K}_j^T\boldsymbol{g}_j & \boldsymbol{K}_j^T\boldsymbol{K}_j \end{pmatrix} \\ &= \hat{\sigma}_j^2\boldsymbol{S}_j^{-1} \end{aligned} \tag{23}$$

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Assuming $\boldsymbol{g}$ and $\boldsymbol{E}$ are independent, we have

$$\boldsymbol{g}_i^T \boldsymbol{K}_j = r_{ij} \sqrt{\boldsymbol{g}_i^T \boldsymbol{K}_i \boldsymbol{g}_j^T \boldsymbol{K}_j} \tag{24}$$

$$\boldsymbol{K}_i^T \boldsymbol{K}_j = r_{ij} \sqrt{\boldsymbol{K}_i^T \boldsymbol{K}_i \boldsymbol{K}_j^T \boldsymbol{K}_j} \tag{25}$$

where r_{ij} is the LD between variant i and j. Therefore, combining equation (23) and LD matrix from reference panel or in-sample LD matrix, sufficient statistic $\boldsymbol{Z}^T\boldsymbol{Z}$ can be constructed and its semi-positive definiteness is guaranteed.

Details in IBSS-ss in SuSiEgxe

The evidence lower bound (ELBO) of SuSiE model equations (1-7) can be obtained by maximizing the ELBO for the SER model l and its ELBO is defined as

$$F_l(q_l, g_l, \sigma^2; \bar{\boldsymbol{r}}_l) = -\frac{N}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} E_{q_l}\left[\|\bar{\boldsymbol{r}}_l - \boldsymbol{\nu}_l\|^2\right] + E_{q_l}\left[log \frac{g_l(\boldsymbol{\nu}_l)}{q_l(\boldsymbol{\nu}_l)}\right] \tag{26}$$

where $\boldsymbol{\nu}_l = \boldsymbol{Z}\boldsymbol{\zeta}_l$, $\boldsymbol{\zeta}_l \sim g_l$, g_l is a pre-specified prior distribution for $\boldsymbol{\zeta}_l$, $q_l \in \mathcal{Q}$ and $\mathcal{Q}$ is all possible distribution on $\boldsymbol{\zeta}$. $\bar{\boldsymbol{r}}_l := E_{q_l}[\boldsymbol{r}_l] := E_{q_l}[\boldsymbol{Y} - \sum_{l' \neq l} \boldsymbol{\nu}_{l'}] = \boldsymbol{Y} - \sum_{l' \neq l} \bar{\boldsymbol{\nu}}_{l'}$ is expected residuals, where $\bar{\boldsymbol{\nu}}_l := (\bar{\nu}_{l1}, \ldots, \bar{\nu}_{ln})^T = E_{q_l}(\nu_{l1}, \ldots, \nu_{ln})^T$ and $\bar{\boldsymbol{\nu}} := (\bar{\boldsymbol{\nu}}_1, \ldots, \bar{\boldsymbol{\nu}}_L)$ are expected value under q_l.

The second term of equation (26) is the expected residual sum of squares (ERSS) under the variational approximation q_l, which is derived as

$$\begin{aligned} E_{q_l}\left[\|\bar{\boldsymbol{r}}_l - \boldsymbol{\nu}_l\|^2\right] &= E_{q_l}\left[\left\|\boldsymbol{Y} - \sum_{l' \neq l} \bar{\boldsymbol{\nu}}_{l'} - \boldsymbol{\nu}_l\right\|^2\right] \\ &= \boldsymbol{Y}^T\boldsymbol{Y} - \sum_{l=1}^{L} 2\,\boldsymbol{Y}^T \bar{\boldsymbol{\nu}}_l + E_{q_l}\left[\left\|\sum_{l' \neq l} \bar{\boldsymbol{\nu}}_{l'} + \boldsymbol{\nu}_l\right\|^2\right] \\ &= \boldsymbol{Y}^T\boldsymbol{Y} - \sum_{l=1}^{L} 2\,\boldsymbol{Y}^T \bar{\boldsymbol{\nu}}_l + \sum_{l=1}^{L} \bar{\boldsymbol{\nu}}_l^T\,\bar{\boldsymbol{\nu}}_l - \bar{\boldsymbol{\nu}}_l^T \bar{\boldsymbol{\nu}}_l + E_{q_l}\left(\boldsymbol{\nu}_l^T \boldsymbol{\nu}_l\right) \end{aligned} \tag{27}$$

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

where $E\left[\boldsymbol{\nu}_{l'}^T \boldsymbol{\nu}_l\right] = 0$ since the posterior mean for different l are independent to each other. $\bar{\boldsymbol{\nu}}_l = \sum_{j=1}^{p} \boldsymbol{Z}_j \, E_{q_l}[\zeta_{lj}]$, and $\bar{\zeta}_{lj} = E_{q_l}[\zeta_{lj}] = \alpha_j \boldsymbol{\mu}_{1,lj}$ is the first moment of ζ_{lj}. For the last term of equation (27),

$$\begin{aligned} \boldsymbol{\nu}_l^T \boldsymbol{\nu}_l &= \sum_{i=1}^{N} \nu_{li}^2 \\ &= \sum_{i=1}^{N} \left(\sum_{j=1}^{p} g_{ij} \, \beta_{lj} + \sum_{j=1}^{p} K_{ij} \, \gamma_{lj} \right)^2 \\ &= \sum_{i=1}^{N} \sum_{j=1}^{p} \left(g_{ij}^2 \beta_{lj}^2 + K_{ij}^2 \gamma_{lj}^2 + 2 g_{ij} K_{ij} \beta_{lj} \gamma_{lj} \right) \end{aligned} \tag{28}$$

where $\beta_{lj}\beta_{lj'} = 0$, $\beta_{lj}\gamma_{lj'} = 0$ and $\gamma_{lj}\gamma_{lj'} = 0$ because only one element in $\boldsymbol{\zeta}_l$ is non-zero under SER model l. Thus,

$$\begin{aligned} E_{q_l}\left(\boldsymbol{\nu}_l^T \boldsymbol{\nu}_l\right) &= \sum_{i=1}^{N} \sum_{j=1}^{p} \left(g_{ij}^2 E_{q_l}[\beta_{lj}^2] + K_{ij}^2 E_{q_l}[\gamma_{lj}^2] + 2 g_{ij} K_{ij} E_{q_l}[\beta_{lj}\gamma_{lj}] \right) \\ &= \sum_{j=1}^{p} \left(\boldsymbol{g}_j^T \boldsymbol{g}_j E_{q_l}[\beta_{lj}^2] + \boldsymbol{K}_j^T \boldsymbol{K}_j E_{q_l}[\gamma_{lj}^2] + 2 \boldsymbol{g}_j^T \boldsymbol{K}_j E_{q_l}[\beta_{lj}\gamma_{lj}] \right) \end{aligned} \tag{29}$$

where the expectation terms are the second moment of ζ_{lj}

$$\begin{aligned} E_{q_l}[\zeta_{lj}\zeta_{lj}^T] &= \begin{bmatrix} E_{q_l}[\beta_{lj}^2] & E_{q_l}[\beta_{lj}\gamma_{lj}] \\ E_{q_l}[\beta_{lj}\gamma_{lj}] & E_{q_l}[\gamma_{lj}^2] \end{bmatrix} \\ &= \alpha_j \left(\boldsymbol{\mu}_{1,lj} \boldsymbol{\mu}_{1,lj}^T + \boldsymbol{\Sigma}_{1,lj} \right), \end{aligned} \tag{30}$$

and $\boldsymbol{Z}_j^T \boldsymbol{Z}_j$ same as in equation (23)

$$\boldsymbol{Z}_j^T \boldsymbol{Z}_j = \begin{bmatrix} \boldsymbol{g}_j^T \boldsymbol{g}_j & \boldsymbol{g}_j^T \boldsymbol{K}_j \\ \boldsymbol{g}_j^T \boldsymbol{K}_j & \boldsymbol{K}_j^T \boldsymbol{K}_j \end{bmatrix}.$$

The third term of equation (26) is $-D_{KL}(q_l \parallel g_l)$, where $D_{KL}(q_l \parallel g_l)$ is Kullback–Leibler (KL) divergence from q_l to g_l.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

$$E_{q_l}\left[log\frac{g_l(\boldsymbol{\nu}_l)}{q_l(\boldsymbol{\nu}_l)}\right] = \sum_{j=1}^{p}\alpha_j\, log\frac{\pi_j}{\alpha_j} + \frac{1}{2\sigma^2}\left(-2E_{q_l}(\boldsymbol{\nu}_l) + E_{q_l}\left(\boldsymbol{\nu}_l^T\boldsymbol{\nu}_l\right)\right) + constant$$

The second term is posterior expected log-likelihood for SER model l.

The IBSS-ss algorithm for SuSiEgxe can be found in Algorithm 1.

Algorithm 1 IBSS-ss for Gene-environment interaction studies.

Require: Sufficient statistics $\boldsymbol{Z}_j^\top\boldsymbol{Z}_j, \boldsymbol{Z}_j^\top\boldsymbol{Y}$, LD matrix from reference panel $\boldsymbol{R}$.

Require: Number of effects, L; initial estimates of hyperparameters $\sigma^2, \boldsymbol{\Sigma}_0$.

Require: Initial estimates of the posterior mean single effects, $\bar{\boldsymbol{\zeta}}_l$, for $l = 1, \ldots, L$.

repeat

step 1: Compute expected residuals.

$\bar{\boldsymbol{\rho}} \leftarrow \boldsymbol{Z}^\top\boldsymbol{Y} - \boldsymbol{Z}^\top\boldsymbol{Z}\sum_{l=1}^{L}\bar{\boldsymbol{\zeta}}_l$

step 2: Fit L SER models iteratively.

for l in $1, \ldots, L$ do

$\bar{\boldsymbol{\rho}}_l \leftarrow \bar{\boldsymbol{\rho}} + \boldsymbol{Z}^\top\boldsymbol{Z}\bar{\boldsymbol{\zeta}}_l$ Disregard lth single effect in residuals.

$\left(\boldsymbol{\alpha}_l, \boldsymbol{\mu}_{1,l}, \boldsymbol{\sigma}_{1,l}^2\right) \leftarrow SER - ss(\boldsymbol{Z}^\top\boldsymbol{Z}, \bar{\boldsymbol{\rho}}_l; \sigma^2, \boldsymbol{\Sigma}_{0l})$ Fit SER to residuals.

$E_{q_l}[\zeta_{lj}] \leftarrow \alpha_j\boldsymbol{\mu}_{1,lj}$ Compute posterior first moment.

$E_{q_l}[\zeta_{lj}\zeta_{lj}^T] \leftarrow \alpha_j\left(\boldsymbol{\mu}_{1,lj}\boldsymbol{\mu}_{1,lj}^T + \boldsymbol{\Sigma}_{1,lj}\right)$ Compute posterior second moments.

$\bar{\boldsymbol{\rho}} \leftarrow \bar{\boldsymbol{\rho}}_l - \boldsymbol{Z}^\top\boldsymbol{Z}\bar{\boldsymbol{\zeta}}_l$ Update expected residuals.

end for

until convergence criterion is satisfied

return $\boldsymbol{\alpha}_1, \mu_{1,1}, \boldsymbol{\Sigma}_{1,1}, \ldots, \boldsymbol{\alpha}_L, \mu_{1,L}, \boldsymbol{\Sigma}_{1,L}$.

medRxiv preprint doi: https://doi.org/10.1101/2025.10.06.25337440; this version posted October 7, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Figure 1. Two-degree-of-freedom fine mapping of locus with lead variant 2:26692756:T_G (rs12476527) on gene *KCNK3*. a) East Asian ancestry. **b**) Cross-population meta-analysis. The figure shows causal SNPs (triangles) on gene *KCNK3* for variant-CURSMK interaction on DBP. Lead variant is indicated by squares. The grey dashed horizontal line marks the threshold for genome-wide significance at $-\log_{10} 5 \times 10^{-9}$. The red dashed vertical lines mark the range of credible sets.

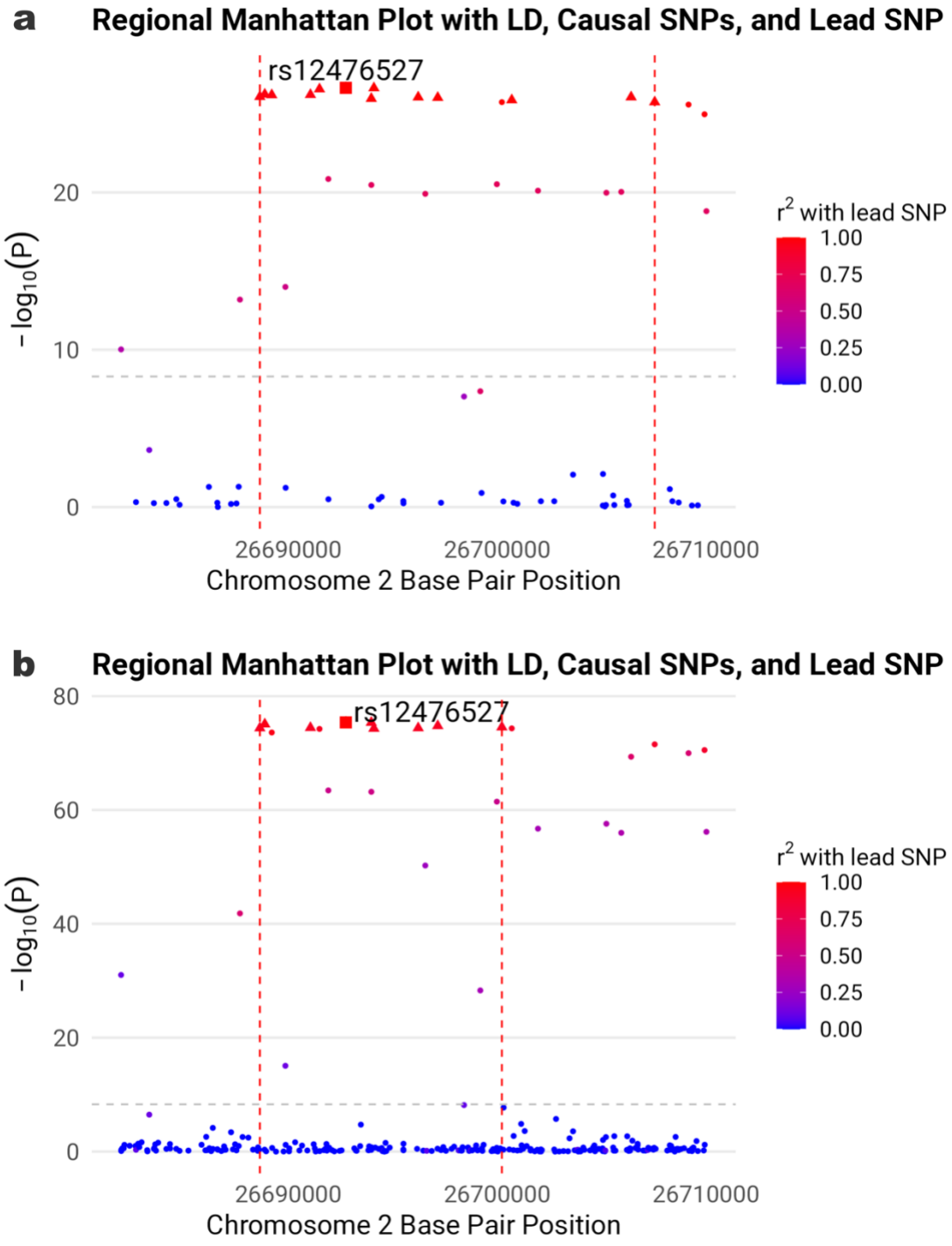


Table1. Novel BP Loci Identified in Cross-Population Meta-Analysis

Locus	Trait	Exposure	Sex	rsID	Position (hg38)	Alleles(E/A)	EAF[a]	N	Closest Gene	Function	Interaction Effect Est	Interaction Effect SE	P_{GxE} or P_{2df}	HetPVal[b]
Novel loci identified by 1df test in combined sex analyses														
1	SBP	CURSMK	COMBINED	rs61840542	10:48289809	T/C	0.0512	888016	*FRMPD2, RP11-542E6.3*	Intergenic, Intergenic	0.0041	0.0007	4.69E-09	0.6639
2	PP	CPD	COMBINED	rs35975869	2:6552172	C/G	0.0101	82854	*AC021021.2, hsa-mir-7515*	Intergenic, Intergenic	0.2553	0.0408	4.087E-10	0.1708
Novel loci identified by 2df test in combined sex analyses														
3	DBP	CURSMK	COMBINED	rs1106090	2:57841606	A/G	0.5935	1179600	*CTD-2026C7.1, VRK2*	Intergenic, Intergenic	0.05627	0.04113	2.822E-09	0.338
4	DBP	CURSMK	COMBINED	rs11293017	6:63509385	A/AT	0.4629	1059650	*EEF1B2P5*	Intergenic	-0.04244	0.04548	2.52E-09	0.1325
5	DBP	CURSMK	COMBINED	–	9:95507136	CG/C	0.777	1056190	*RP11-435O5.5*	ncRNA_exonic	-0.02874	0.05344	3.398E-09	0.2432
6	DBP	CURSMK	COMBINED	rs8070260	17:36511076	A/G	0.4695	1179170	*MYO19*	Intronic	0.06636	0.04069	1.358E-09	0.2403
7	DBP	CURSMK	COMBINED	rs6123158	20:52337620	A/G	0.5765	1182960	*RP4-723E3.1*	ncRNA_intronic	0.003827	0.04213	2.019E-09	0.5572
8	SBP	CURSMK	COMBINED	rs140470470	4:139295340	T/C	0.0152	1074210	*NDUFC1*	Intronic	0.1008	0.2772	1.187E-10	0.3379
9	PP	CURSMK	COMBINED	rs34408570	3:33200699	C/G	0.116	1170410	*SUSD5*	Intronic	-0.1714	0.07414	1.958E-09	0.05021
10	PP	CURSMK	COMBINED	rs111563399	3:187170275	T/C	0.9472	900381	*RPL39L*	Intronic	0.04153	0.1222	1.29E-09	0.8205
11	PP	CPD	COMBINED	rs143429637	12:59947323	A/G	0.0076	92630	*RP11-813P10.2, AC091517.1*	Intergenic, Intergenic	0.05556	0.0306	1.19E-14	0.03971
Novel loci identified by 2df test in Males														
12	DBP	CURSMK	MALE	rs72832924	17:39740603	T/C	0.1624	648750	*GRB7*	Intronic	-0.1119	0.07346	2.767E-10	0.07679

Note: a) Effect allele frequency. b) P values of testing sex heterogeneity.

Table 2. BP Known Loci Identified Using the 1df Interaction Model.

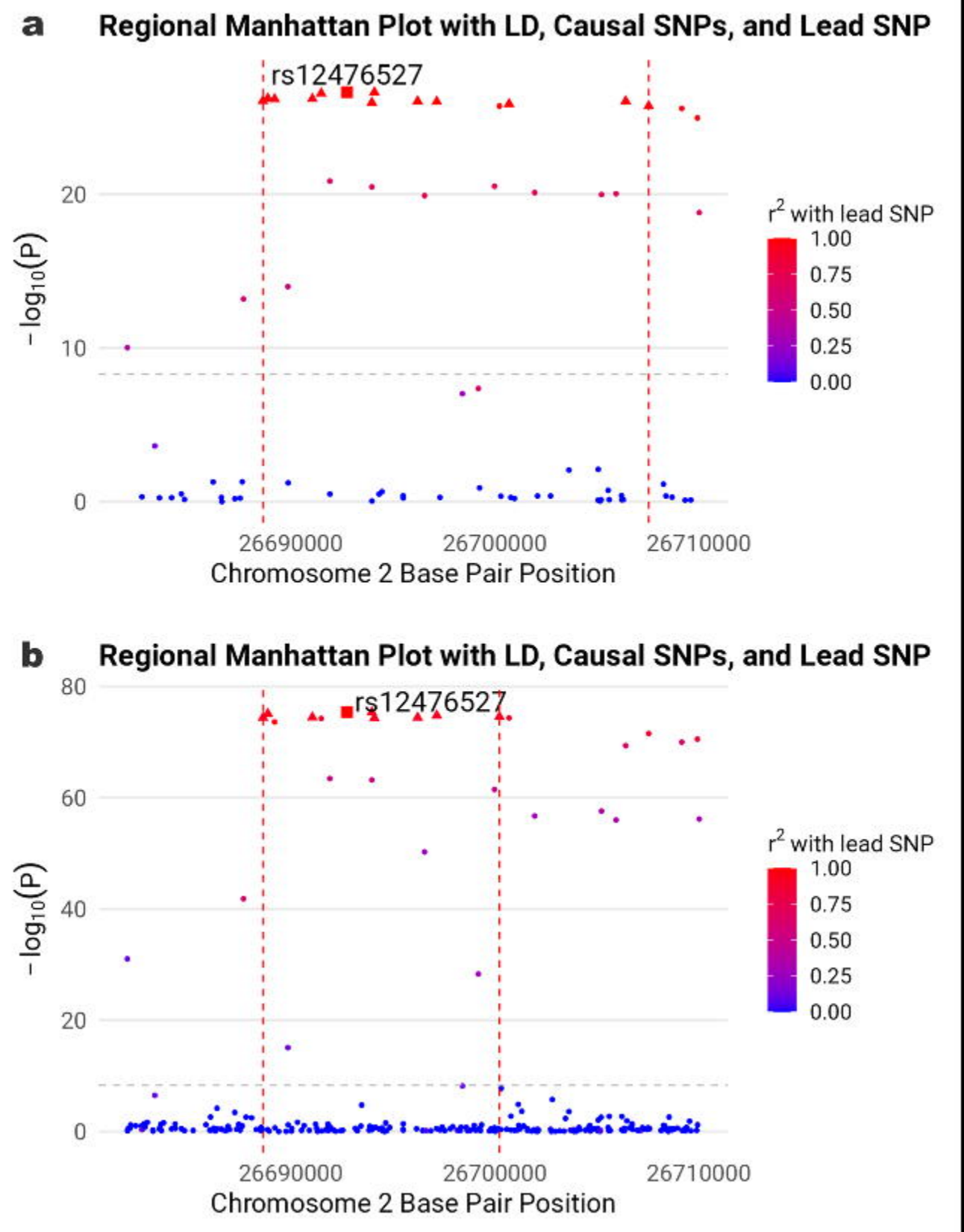
Locus	Population	Trait	Exposure	Sex	rsID	Position (hg38)	Alleles (E/A)	EAF[a]	N	Closest Gene	Nearby Known BP Protein-coding Genes within 500kb	Interaction Effect	Interaction Effect SE	P_{GxE} or P_{GxE,m_G}	HetPVal[c]
1	EUR	PP	PY	FEMALE	rs56654546	1:40091270	A/T	0.8953	32039	*PPT1*	*COL9A2*	0.0305	0.0092	9.646E-4[b]	0.8835
2	EAS	DBP	CURSMK	MALE	rs11065783	12:110958445	A/G	0.7299	120049	*LINC01404*	*PPTC7;MYL2; CUX2;SH2B3*	0.5532	0.1319	2.718E-5[b]	0.3930
3	EAS	SBP	CURSMK	MALE	rs78572043	12:111080689	A/G	0.8037	118774	*CUX2*	*MYL2;CUX2; SH2B3;ATXN2*	0.9466	0.2329	4.81E-5[b]	0.6556
4	EAS	PP	CURSMK	COMBINED	rs3782886	12:111672685	T/C	0.7664	246722	*BRAP*	*CUX2;SH2B3;ATXN2;ACAD10;ALDH2;ERP29; NAA25;TRAFD1*	0.8102	0.1201	1.496E-11	0.8834
		SBP						0.7616	246817			1.833	0.1819	6.982E-24	0.3451
5	EAS	DBP	CURSMK	FEMALE	rs116873087	12:112074109	C/G	0.2303	121133	*NAA25*	*ACAD10;ALDH2;ERP29; NAA25;TRAFD1; HECTD4;RPL6-ALDH1*	-1.3513	0.2676	4.422E-07[b]	0.7372
				COMBINED				0.2232	246773			-1.0736	0.121	6.979E-19	0.1518
6	EAS	DBP	CURSMK	MALE	rs12231737	12:112136812	T/C	0.2452	120049	*TRAFD1*	*ACAD10;ALDH2;ERP29; NAA25;TRAFD1; HECTD4;RPL6-ALDH1*	-0.5786	0.138	2.758E-5[b]	0.5172
		SBP						0.2429	120074			-0.9374	0.2149	1.285E-5[b]	0.4551
7	CPMA	DBP	CURSMK	COMBINED	rs11066280	12:112379979	A/T	0.2447	281269	*HECTD4*	*ERP29;NAA25;TRAFD1; HECTD4;RPL6-ALDH1*	-0.9517	0.1166	3.319E-16	0.1495
		PP						0.2383	281218			-0.7257	0.1199	1.43E-09	0.5824
		SBP						0.2433	281313			-1.6825	0.1817	2.07E-20	0.2010
8	CPMA	PP	CURSMK	COMBINED	rs4237919	12:20091096	A/G	0.3756	1170940	*LINC02398*	*PDE3A*	-0.2266	0.0484	2.869E-06[b]	0.5847
9	EAS	SBP	CURSMK	MALE	rs1327235	20:10988382	A/G	0.4894	120074	*LINC02871*	*SLX4IP; JAG1;BANF2*	0.7364	0.169	1.32E-5[b]	0.4131
10	EUR	PP	PY	COMBINED	rs16892639	8:119450019	A/T	0.257	88822	-	*MAL2;SNORA32;NOV; CCN3 - ENPP2; ENPP2;DEPTOR*	0.0145	0.0038	1.223E-4[b]	0.0860

Note: a) Effect allele frequency. b) Bonferroni corrected significant loci at 1df interaction test: $P_{GxE,m_G} \leq 0.05/m_G$ and $FDR_{GxE,m_G} < 0.05$, where FDR_{GxE,m_G} is 1df GxE false discovery rate (FDR) calculated based on m_G variants, and m_G represents the number of variants with marginal test $P_G \leq 10^{-5}$ from Model 2. c) P values of testing sex heterogeneity.

Table 3. Smoking Status-Stratified Summary Statistics at Significant 1df Interaction Loci Known for BP

Locus[a]	Population	Trait	Exposure	Sex	rsID	Position (hg38)	Alleles (E/A)	EAF[b]	N	N (E = 1)[a]	Genetic Effect (E = 1)	Genetic Effect SE (E = 1)	P (E = 1)	N (E = 0)[b]	Genetic Effect (E = 0)	Genetic Effect SE (E = 0)	P (E = 0)
2	EAS	DBP	CURSMK	MALE	rs11065783	12:110958445	A/G	0.7299	120049	32416	1.101	0.114	3.793E-22	87632	0.527	0.068	1.271E-14
3	EAS	SBP	CURSMK	MALE	rs78572043	12:111080689	A/G	0.8037	118774	32079	2.779	0.204	3.595E-42	86694	1.771	0.120	1.570E-49
4	EAS	PP	CURSMK	COMBINED	rs3782886	12:111672685	T/C	0.7664	246722	40820	1.087	0.113	**6.855E-22**	205901	0.268	0.049	**3.917E-08**
		SBP						0.7616	246817	40843	2.727	0.172	2.478E-56	205973	0.886	0.075	6.575E-32
5	EAS	DBP	CURSMK	FEMALE	rs116873087	12:112074109	C/G	0.2303	121133	8196	-1.584	0.256	**6.204E-10**	112936	-0.240	0.067	**3.428E-04**
				COMBINED				0.2232	246773	40828	-1.692	0.114	2.923E-50	205944	-0.623	0.049	1.477E-36
6	EAS	DBP	CURSMK	MALE	rs12231737	12:112136812	T/C	0.2452	120049	8196	-1.462	0.249	**4.186E-09**	112936	-0.251	0.066	**1.299E-04**
		SBP						0.2429	120074	32430	-2.889	0.186	3.749E-54	87643	-1.854	0.112	7.202E-62
7	CPMA	DBP	CURSMK	COMBINED	rs11066280	12:112379979	A/T	0.2447	281269	53924	-1.511	0.111	3.142E-42	227344	-0.567	0.048	1.338E-31
		PP						0.2383	281218	53915	-0.993	0.114	**3.251E-18**	227302	-0.257	0.049	**1.669E-07**
		SBP						0.2433	281313	53935	-2.528	0.175	2.439E-47	227377	-0.837	0.076	5.785E-28
8	CPMA	PP	CURSMK	COMBINED	rs4237919	12:20091096	A/G	0.3756	1170940	224496	-0.093	0.046	**0.042**	946443	0.129	0.020	**2.878E-10**
9	EAS	SBP	CURSMK	MALE	rs1327235	20:10988382	A/G	0.4894	120074	32430	0.097	0.145	**0.505**	87643	-0.618	0.089	**4.897E-12**

Note: a) Same locus has the same number same as in Table 2. b) E = 1 is exposed group to smoking. c) E = 0 is unexposed group to smoking


a
Regional Manhattan Plot with LD, Causal SNPs, and Lead SNP
rs12476527
20
10
0
$-\log_{10}(P)$
r^2 with lead SNP
1.00
0.75
0.50
0.25
0.00
26690000
26700000
26710000
Chromosome 2 Base Pair Position
b
Regional Manhattan Plot with LD, Causal SNPs, and Lead SNP
rs12476527
80
60
40
20
0
$-\log_{10}(P)$
r^2 with lead SNP
1.00
0.75
0.50
0.25
0.00
26690000
26700000
26710000
Chromosome 2 Base Pair Position