1 Genetic analysis of over one million people identifies 535 new loci associated with blood

2 pressure traits.

- 4 Short title: blood pressure GWAS in one million people
- 5 Evangelos Evangelou^{1,2*}, Helen R Warren^{3,4*}, David Mosen-Ansorena^{1*}, Borbala Mifsud^{3*}, Raha
- 6 Pazoki^{1*}, He Gao^{1,5*}, Georgios Ntritsos^{2*}, Niki Dimou^{2*}, Claudia P Cabrera^{3,4}, Ibrahim Karaman¹,
- Fu Liang Ng³, Marina Evangelou¹,6, Katarzyna Witkowska³, Evan Tzanis³, Jacklyn N Hellwege²,
- 8 Ayush Giri⁸, Digna R Velez Edwards⁸, Yan V Sun^{9,10}, Kelly Cho^{11,12}, J.Michael Gaziano^{11,12}, Peter
- 9 WF Wilson¹³, Philip S Tsao¹⁴, Csaba P Kovesdy¹⁵, Tonu Esko^{16,17}, Reedik Mägi¹⁶, Lili Milani¹⁶,
- 10 Peter Almgren¹⁸, Thibaud Boutin¹⁹, Stéphanie Debette^{20,21}, Jun Ding²², Franco Giulianini²³,
- Elizabeth G Holliday²⁴, Anne U Jackson²⁵, Ruifang Li-Gao²⁶, Wei-Yu Lin²⁷, Jian'an Luan²⁸, Massimo
- 12 Mangino^{29,30}, Christopher Oldmeadow²⁴, Bram Peter Prins³¹, Yong Qian²², Muralidharan
- 13 Sargurupremraj²¹, Nabi Shah^{32,33}, Praveen Surendran²⁷, Sébastien Thériault^{34,35}, Niek
- 14 Verweij^{17,36,37}, Sara M Willems²⁸, Jing-Hua Zhao²⁸, Philippe Amouyel³⁸, John Connell³⁹, Renée de
- 15 Mutsert²⁶, Alex SF Doney³², Martin Farrall^{40,41}, Cristina Menni²⁹, Andrew D Morris⁴², Raymond
- 16 Noordam⁴³, Guillaume Paré³⁴, Neil R Poulter⁴⁴, Denis C Shields⁴⁵, Alice Stanton⁴⁶, Simon Thom⁴⁷,
- Gonçalo Abecasis⁴⁸, Najaf Amin⁴⁹, Dan E Arking⁵⁰, Kristin L Ayers^{51,52}, Caterina M Barbieri⁵³,
- Chiara Batini⁵⁴, Joshua C Bis⁵⁵, Tineka Blake⁵⁴, Murielle Bochud⁵⁶, Michael Boehnke²⁵, Eric
- 19 Boerwinkle⁵⁷, Dorret I Boomsma⁵⁸, Erwin P Bottinger⁵⁹, Peter S Braund^{60,61}, Marco Brumat⁶²,
- 20 Archie Campbell^{63,64}, Harry Campbell⁶⁵, Aravinda Chakravarti⁵⁰, John C Chambers^{1,5,66-68}, Ganesh
- 21 Chauhan⁶⁹, Marina Ciullo^{70,71}, Massimiliano Cocca⁷², Francis Collins⁷³, Heather J Cordell⁵¹, Gail
- Davies^{74,75}, Martin H de Borst⁷⁶, Eco J de Geus⁵⁸, Ian J Deary^{74,75}, Joris Deelen⁷⁷, Fabiola Del Greco
- 23 M⁷⁸, Cumhur Yusuf Demirkale⁷⁹, Marcus Dörr^{80,81}, Georg B Ehret^{50,82}, Roberto Elosua^{83,84}, Stefan
- 24 Enroth⁸⁵, A Mesut Erzurumluoglu⁵⁴, Teresa Ferreira^{86,87}, Mattias Frånberg⁸⁸⁻⁹⁰, Oscar H Franco⁹¹,
- 25 Ilaria Gandin⁶², Paolo Gasparini^{62,72}, Vilmantas Giedraitis⁹², Christian Gieger⁹³⁻⁹⁵, Giorgia
- 26 Girotto^{62,72}, Anuj Goel^{40,41}, Alan J Gow^{74,96}, Vilmundur Gudnason^{97,98}, Xiuqing Guo⁹⁹, Ulf
- 27 Gyllensten⁸⁵, Anders Hamsten^{88,89}, Tamara B Harris¹⁰⁰, Sarah E Harris^{63,74}, Catharina A
- Hartman¹⁰¹, Aki S Havulinna^{102,103}, Andrew A Hicks⁷⁸, Edith Hofer^{104,105}, Albert Hofman^{91,106},
- Jouke-Jan Hottenga⁵⁸, Jennifer E Huffman^{19,107,108}, Shih-Jen Hwang^{107,108}, Erik Ingelsson^{109,110},
- 30 Alan James^{111,112}, Rick Jansen¹¹³, Marjo-Riitta Jarvelin^{1,5,114-116}, Roby Joehanes^{107,117}, Åsa
- 31 Johansson⁸⁵, Andrew D Johnson^{107,118}, Peter K Joshi⁶⁵, Pekka Jousilahti¹⁰², J Wouter Jukema¹¹⁹,
- 32 Antti Jula¹⁰², Mika Kähönen^{120,121}, Sekar Kathiresan^{17,36,122}, Bernard D Keavney^{123,124}, Kay-Tee
- 33 Khaw¹²⁵, Paul Knekt¹⁰², Joanne Knight¹²⁶, Ivana Kolcic¹²⁷, Jaspal S Kooner^{5,67,68,128}, Seppo
- 34 Koskinen¹⁰², Kati Kristiansson¹⁰², Zoltan Kutalik^{56,129}, Maris Laan¹³⁰, Marty Larson¹⁰⁷, Lenore J
- Launer¹⁰⁰, Benjamin Lehne¹, Terho Lehtimäki^{131,132}, David CM Liewald^{74,75}, Li Lin⁸², Lars Lind¹³³,
- Cecilia M Lindgren^{40,87,134}, YongMei Liu¹³⁵, Ruth JF Loos^{28,59,136}, Lorna M Lopez^{74,137,138}, Yingchang
- 37 Lu⁵⁹, Leo-Pekka Lyytikäinen^{131,132}, Anubha Mahajan⁴⁰, Chrysovalanto Mamasoula¹³⁹, Jaume
- 38 Marrugat⁸³, Jonathan Marten¹⁹, Yuri Milaneschi¹⁴⁰, Anna Morgan⁶², Andrew P Morris^{40,141},
- 39 Alanna C Morrison¹⁴², Peter I Munson⁷⁹, Mike A Nalls^{143,144}, Priyanka Nandakumar⁵⁰, Christopher
- 40 P Nelson^{60,61}, Teemu Niiranen^{102,145}, Ilja M Nolte¹⁴⁶, Teresa Nutile⁷⁰, Albertine J Oldehinkel¹⁴⁷,
- Ben A Oostra⁴⁹, Paul F O'Reilly¹⁴⁸, Elin Org¹⁶, Sandosh Padmanabhan^{64,149}, Walter Palmas¹⁵⁰,
- 42 Aarno Palotie^{103,151,152}, Alison Pattie⁷⁵, Brenda WJH Penninx¹⁴⁰, Markus Perola^{102,103,153}, Annette
- 43 Peters^{94,95,154}, Ozren Polasek^{127,155}, Peter P Pramstaller^{78,156,157}, Quang Tri Nguyen⁷⁹, Olli T
- 44 Raitakari^{158,159}, Meixia Ren¹⁶⁰, Rainer Rettig¹⁶¹, Kenneth Rice¹⁶², Paul M Ridker^{23,163}, Janina S
- 45 Ried⁹⁴, Harriëtte Riese¹⁴⁷, Samuli Ripatti^{103,164}, Antonietta Robino⁷², Lynda M Rose²³, Jerome I
- 46 Rotter⁹⁹, Igor Rudan¹⁶⁵, Daniela Ruggiero^{70,71}, Yasaman Saba¹⁶⁶, Cinzia F Sala⁵³, Veikko
- 47 Salomaa¹⁰², Nilesh J Samani^{60,61}, Antti-Pekka Sarin¹⁰³, Reinhold Schmidt¹⁰⁴, Helena Schmidt¹⁶⁶,

- 48 Nick Shrine⁵⁴, David Siscovick¹⁶⁷, Albert V Smith^{97,98}, Harold Snieder¹⁴⁶, Siim Sõber¹³⁰, Rossella
- 49 Sorice⁷⁰, John M Starr^{74,168}, David J Stott¹⁶⁹, David P Strachan¹⁷⁰, Rona J Strawbridge^{88,89}, Johan
- 50 Sundström¹³³, Morris A Swertz¹⁷¹, Kent D Taylor⁹⁹, Alexander Teumer^{81,172}, Martin D Tobin⁵⁴,
- Maciej Tomaszewski^{123,124}, Daniela Toniolo⁵³, Michela Traglia⁵³, Stella Trompet^{119,173}, Jaakko
- 52 Tuomilehto¹⁷⁴⁻¹⁷⁷, Christophe Tzourio²¹, André G Uitterlinden^{91,178}, Ahmad Vaez^{146,179}, Peter J van
- der Most¹⁴⁶, Cornelia M van Duijn⁴⁹, Anne-Claire Vergnaud¹, Germaine C Verwoert⁹¹, Veronique
- Vitart¹⁹, Uwe Völker^{81,180}, Peter Vollenweider¹⁸¹, Dragana Vuckovic^{62,182}, Hugh Watkins^{40,41},
- 55 Sarah H Wild¹⁸³, Gonneke Willemsen⁵⁸, James F Wilson^{19,65}, Alan F Wright¹⁹, Jie Yao⁹⁹, Tatijana
- Zemunik¹⁸⁴, Weihua Zhang^{1,67}, John R Attia²⁴, Adam S Butterworth^{27,185}, Daniel I Chasman^{23,163},
- David Conen^{186,187}, Francesco Cucca^{188,189}, John Danesh^{27,185}, Caroline Hayward¹⁹, Joanna MM
- 58 Howson²⁷, Markku Laakso¹⁹⁰, Edward G Lakatta¹⁹¹, Claudia Langenberg²⁸, Olle Melander¹⁸,
- Dennis O Mook-Kanamori^{26,192}, Colin NA Palmer³², Lorenz Risch¹⁹³⁻¹⁹⁵, Robert A Scott²⁸, Rodney J
- 60 Scott²⁴, Peter Sever¹²⁸, Tim D Spector²⁹, Pim van der Harst¹⁹⁶, Nicholas J Wareham²⁸, Eleftheria
- 61 Zeggini³¹, Daniel Levy^{107,118}, Patricia B Munroe^{3,4}, Christopher Newton-Cheh^{134,197,198}, Morris J
- 62 Brown^{3,4}, Andres Metspalu¹⁶, Adriana M Hung¹⁹⁹, Christopher J O'Donnell²⁰⁰, Todd L Edwards⁷
- on behalf of the Million Veteran Program, Bruce M. Psaty^{201,202}, Ioanna Tzoulaki^{1,2,5*}, Michael R
- Barnes^{3,4*}, Louise V Wain^{54*}, Paul Elliott^{1,5,203-205*}‡, Mark J Caulfield^{3,4*}‡

- * Equal contribution
- 4 Corresponding authors

- 69 1. Department of Epidemiology and Biostatistics, Imperial College London, London, UK.
- Department of Hygiene and Epidemiology, University of Ioannina Medical
 School, Ioannina, Greece.
- William Harvey Research Institute, Barts and The London School of Medicine and
 Dentistry, Queen Mary University of London, London, UK.
- 75 4. National Institute for Health Research, Barts Cardiovascular Biomedical Research Center, Queen Mary University of London, London, UK.
- 77 5. MRC-PHE Centre for Environment and Health, Imperial College London, London, UK.
- 79 6. Department of Mathematics, Imperial College London, London, UK
- 7. Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical
- 82 Center, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA.
- 84 8. Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Department of 85 Obstetrics and Gynecology, Vanderbilt University Medical Center; Tennessee 86 Valley Health Systems VA, Nashville, TN, USA.
- 9. Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, USA.
- Department of Biomedical Informatics, Emory University School of Medicine,Atlanta, GA, USA.
- 91 11. Massachusetts Veterans Epidemiology Research and Information Center 92 (MAVERIC), VA Boston Healthcare System, Boston, USA.

- Division of Aging, Department of Medicine, Brigham and Women's Hospital,
 Boston, MA, Department of Medicine, Harvard Medical School, Boston, MA, USA.
- 95 13. Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA, USA.
- 97 14. VA Palo Alto Health Care System; Division of Cardiovascular Medicine, Stanford University School of Medicine, CA, USA.
- 99 15. Nephrology Section, Memphis VA Medical Center and University of Tennessee 100 Health Science Center, Memphis, TN, USA.
- 101 16. Estonian Genome Center, University of Tartu, Tartu, Estonia.
- 17. Program in Medical and Population Genetics, Broad Institute of Harvard and
 MIT, Cambridge, MA, USA.
- 104 18. Department Clinical Sciences, Malmö, Lund University, Malmö, Sweden.
- 19. MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine,
 University of Edinburgh, Western General Hospital, Edinburgh, Scotland, UK
- 107 20. Department of Neurology, Bordeaux University Hospital, Bordeaux, France.
- Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, CHU
 Bordeaux, Bordeaux, France.
- 110 22. Laboratory of Genetics and Genomics, NIA/NIH, Baltimore, MD, USA.
- Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA,USA.
- Hunter Medical Reseach Institute and Faculty of Health, University of Newcastle, New Lambton Heights, New South Wales, Australia.
- Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA.
- Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.
- MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and
 Primary Care, University of Cambridge, Cambridge, UK.
- 121 28. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK.
- 123 29. Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK.
- 125 30. NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, 126 London, UK.
- 127 31. Wellcome Trust Sanger Institute, Hinxton, UK.
- 128 32. Division of Molecular and Clinical Medicine, School of Medicine, University of Dundee, UK.
- 130 33. Department of Pharmacy, COMSATS Institute of Information Technology,131 Abbottabad, Pakistan.
- 34. Department of Pathology and Molecular Medicine, McMaster University,
 Hamilton, Canada.
- 134 35. Institut universitaire de cardiologie et de pneumologie de Québec-Université 135 Laval, , Quebec City, Canada.
- 36. Cardiovascular Research Center and Center for Human Genetic Research,
 Massachusetts General Hospital, Boston, Massachusetts, MA, USA.
- 138 37. University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, The Netherlands.

- 140 38. University of Lille, Inserm, Centre Hosp. Univ Lille, Institut Pasteur de Lille,
- 141 UMR1167 RID-AGE Risk factors and molecular determinants of aging-related
 142 diseases, Epidemiology and Public Health Department, Lille, France.
- 143 39. University of Dundee, Ninewells Hospital & Medical School, Dundee, , UK.
- 40. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
- 145 41. Division of Cardiovascular Medicine, Radcliffe Department of Medicine,
 146 University of Oxford, Oxford, UK.
- 42. Usher Institute of Population Health Sciences and Informatics, University of
 Edinburgh, UK.
- 149 43. Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden
 150 University Medical Center, Leiden, The Netherlands.
- 151 44. Imperial Clinical Trials Unit, Stadium House, 68 Wood Lane, London, UK.
- 152 45. School of Medicine, University College Dublin, Ireland.
- 46. Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland,Dublin, Ireland.
- 47. International Centre for Circulatory Health, Imperial College London, London,UK.
- 48. Center for Statistical Genetics, Dept. of Biostatistics, SPH II, Washington Heights,Ann Arbor, MI, USA.
- 49. Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC,Rotterdam, the Netherlands.
- 50. Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic
 Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
- 163 51. Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK.
- 164 52. Sema4, a Mount Sinai venture, Stamford, CT, USA.
- 165 53. Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milano,166 Italy.
- 167 54. Department of Health Sciences, University of Leicester, Leicester, UK.
- 168 55. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA.
- 170 56. Institute of Social and Preventive Medicine, University Hospital of Lausanne,Lausanne, Switzerland.
- Human Genetics Center, School of Public Health, The University of Texas Health
 Science Center at Houston and Human Genome Sequencing Center, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX, USA.
- Department of Biological Psychology, Vrije Universiteit Amsterdam, EMGO+ institute, VU University medical center, Amsterdam, the Netherlands.
- The Charles Bronfman Institute for Personalized Medicine, Icachn School of
 Medicine at Mount Sinai, NY, USA.
- 179 60. Department of Cardiovascular Sciences, University of Leicester, Leicester, UK.
- 180 61. NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Groby Road, Leicester, UK.
- 182 62. Department of Medical, Surgical and Health Sciences, University of Trieste, , Trieste, Italy.
- Medical Genetics Section, Centre for Genomic and Experimental Medicine,
 Institute of Genetics and Molecular Medicine, University of Edinburgh,
 Edinburgh, UK.
- 187 64. Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK.

- 189 65. Centre for Global Health Research, Usher Institute of Population Health Sciences 190 and Informatics, University of Edinburgh, Edinburgh, Scotland, UK
- Lee Kong Chian School of Medicine, Nanyang Technological University,
 Singapore, Singapore.
- 193 67. Department of Cardiology, Ealing Hospital, Middlesex, UK.
- 194 68. Imperial College Healthcare NHS Trust, London, UK.
- 195 69. Centre for Brain Research, Indian Institute of Science, Bangalore, India.
- 196 70. Institute of Genetics and Biophysics "A. Buzzati-Traverso", CNR, Napoli, Italy.
- 197 71. IRCCS Neuromed, Pozzilli, Isernia, Italy.
- 198 72. Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy.
- 73. Medical Genomics and Metabolic Genetics Branch, National Human Genome
 Research Institute, NIH, Bethesda, MD, USA.
- 74. Centre for Cognitive Ageing and Cognitive Epidemiology, University of
 Edinburgh, 7 George Square, Edinburgh, UK.
- 203 75. Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, 204 UK.
- 76. Department of Internal Medicine, Division of Nephrology, University of
 Groningen, University Medical Center Groningen, Groningen, The Netherlands.
- 77. Department of Molecular Epidemiology, Leiden University Medical Center,
 Leiden, the Netherlands.
- 209 78. Institute for Biomedicine, Eurac Research, Bolzano, Italy Affiliated Institute of the University of Lübeck, Lübeck, Germany.
- 211 79. Mathematical and Statistical Computing Laboratory, Office of Intramural
- Research, Center for Information Technology, National Institutes of Health,
 Bethesda, MD, USA.
- 214 80. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany.
- 216 81. DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany.
- 218 82. Cardiology, Department of Medicine, Geneva University Hospital, Geneva, Switzerland.
- 220 83. CIBERCV & Cardiovascular Epidemiology and Genetics, IMIM. Dr Aiguader 88,
 221 Barcelona, Spain.
- 222 84. Faculty of Medicine, Universitat de Vic-Central de Catalunya, Vic, Spain.
- 223 85. Department of Immunology, Genetics and Pathology, Uppsala Universitet, Science for Life Laboratory, Uppsala, Sweden.
- Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, UK.
- Big Data Institute, Li Ka Shing Center for Health Information and Discovery, Oxford University, Old Road, Oxford, UK.
- 229 88. Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet. Stockholm. Sweden.
- 231 89. Centre for Molecular Medicine, L8:03, Karolinska Universitetsjukhuset, Solna, Sweden.
- 233 90. Department of Numerical Analysis and Computer Science, Stockholm University, Stockholm, Sweden.
- 235 91. Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands.
- 236 92. Department of Public Health and Caring Sciences, Geriatrics, Uppsala, Sweden.

- Page 237 Page 238 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
- 94. Institute of Epidemiology, Helmholtz Zentrum München, German Research
 Center for Environmental Health, Neuherberg, Germany.
- 241 95. German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany.
- 96. Department of Psychology, School of Social Sciences, Heriot-Watt University,
 Edinburgh, UK.
- 244 97. Faculty of Medicine, University of Iceland, Reykjavik, Iceland.
- 245 98. Icelandic Heart Association, Kopavogur, Iceland.
- 246 99. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA.
- Intramural Research Program, Laboratory of Epidemiology, Demography, and
 Biometry, National Institute on Aging, Bethesda, MD, USA.
- Department of Psychiatry, University of Groningen, University Medical Center
 Groningen, Groningen, The Netherlands.
- Department of Public Health Solutions, National Institute for Health and Welfare (THL), Helsinki, Finland.
- 254 103. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, 255 Finland.
- 256 104. Clinical Division of Neurogeriatrics, Department of Neurology, Medical
 257 University of Graz, Graz, Austria.
- 105. Institute for Medical Informatics, Statistics and Documentation, Medical
 University of Graz, Graz, Austria.
- 260 106. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
- 107. National Heart, Lung and Blood Institute's Framingham Heart Study,
 Framingham, MA, USA.
- The Population Science Branch, Division of Intramural Research, National Heart
 Lung and Blood Institute national Institute of Health, Bethesda, MD, USA.
- Department of Medical Sciences, Molecular Epidemiology and Science for Life
 Laboratory, Uppsala University, Uppsala, Sweden.
- 268 110. Division of Cardiovascular Medicine, Department of Medicine, Stanford
 269 University School of Medicine, Stanford, CA USA.
- 270 111. Department of Pulmonary Physiology and Sleep, Sir Charles Gairdner Hospital,
 271 Hospital Avenue, Nedlands, Australia.
- 272 112. School of Medicine and Pharmacology, University of Western Australia.
- 273 113. Department of Psychiatry, VU University Medical Center, Amsterdam
 274 Neuroscience, Amsterdam, the Netherlands.
- 275 114. Biocenter Oulu, University of Oulu, Oulu, Finland.
- 276 115. Center For Life-course Health Research, University of Oulu, Oulu Finland.
- 277 116. Unit of Primary Care, Oulu University Hospital, Oulu, Oulu, Finland.
- 278 117. Hebrew SeniorLife, Harvard Medical School, Boston, MA, USA.
- 279 118. Population Sciences Branch, National Heart, Lung and Blood Institute, National
 280 Institutes of Health, Bethesda, MD, USA.
- 281 119. Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands.
- 283 120. Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland.

- Department of Clinical Physiology, Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Life Sciences, University of Tampere,
 Tampere, Finland.
- 288 122. Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA.
- 290 123. Division of Cardiovascular Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK.
- 124. Division of Medicine, Manchester University NHS Foundation Trust, Manchester
 Academic Health Science Centre, Manchester, UK
- Department of Public Health and Primary Care, Institute of Public Health,
 University of Cambridge, Cambridge, UK.
- 296 126. Data Science Institute and Lancaster Medical School, Lancaster, UK.
- 297 127. Department of Public Health, Faculty of Medicine, University of Split, Croatia.
- 298 128. National Heart and Lung Institute, Imperial College London, London, UK.
- 299 129. Swiss Institute of Bioinformatics, Lausanne, Switzerland.
- 300 130. Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, 301 Estonia.
- 302 131. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland.
- 303 132. Department of Clinical Chemistry, Finnish Cardiovascular Research Center 304 Tampere, Faculty of Medicine and Life Sciences, University of Tampere,
 305 Tampere, Finland
- 306 133. Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala 307 University, Uppsala, Sweden.
- 308 134. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, 309 USA.
- 310 135. Division of Public Health Sciences, Wake Forest School of Medicine, Winston-311 Salem, NC, USA.
- 312 136. Mindich Child health Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY, USA.
- Department of Psychiatry, Royal College of Surgeons in Ireland, Education and
 Research Centre, Beaumont Hospital, Dublin, Ireland.
- 138. University College Dublin, UCD Conway Institute, Centre for Proteome Research,
 317 UCD, Belfield, Dublin, Ireland.
- 318 139. Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK.
- 319 140. Department of Psychiatry, Amsterdam Public Health and Amsterdam
- Neuroscience, VU University Medical Center/GGZ inGeest, Amsterdam, The Netherlands.
- 322 141. Department of Biostatistics, University of Liverpool, Block F, Waterhouse Building, Liverpool, UK.
- 324 142. Department of Epidemiology, Human Genetics and Environmental Sciences,
- School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA.
- 327 143. Data Tecnica International, Glen Echo, MD, USA.
- 328 144. Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA.
- 329 145. Department of Medicine, Turku University Hospital and University of Turku, Finland.
- 331 146. Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

- 333 147. Interdisciplinary Center Psychopathology and Emotion regulation (ICPE),
- University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
- 336 148. SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King's337 College London, London, UK.
- 338 149. British Heart Foundation Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.
- 150. Department of Medicine, Columbia University Medical Center, New York, NY,USA.
- Analytic and Translational Genetics Unit, Department of Medicine, Department of
 Neurology and Department of Psychiatry Massachusetts General Hospital,
 Boston, MA, USA.
- The Stanley Center for Psychiatric Research and Program in Medical and
 Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA,
 USA.
- 349 153. University of Tartu, Tartu, Estonia.
- 350 154. German Center for Cardiovascular Disease Research (DZHK), partner site Munich, Neuherberg, Germany.
- 352 155. Psychiatric hospital "Sveti Ivan", Zagreb, Croatia.
- 353 156. Department of Neurology, General Central Hospital, Bolzano, Italy.
- 354 157. Department of Neurology, University of Lübeck, Lübeck, Germany.
- 158. Department of Clinical Physiology and Nuclear Medicine, Turku UniversityHospital, Turku, Finland.
- 357 159. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
- Fujian Key Laboratory of Geriatrics, Department of Geriatric Medicine, Fujian
 Provincial Hospital, Fujian Medical University, Fuzhou, China.
- 361 161. Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany.
- 362 162. Department of Biostatistics University of Washington, Seattle, WA, USA.
- 363 163. Harvard Medical School, Boston MA.
- 364 164. Public health, Faculty of Medicine, University of Helsinki, Finland
- 365 165. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Scotland, UK.
- 367 166. Gottfried Schatz Research Center for Cell Signaling, Metabolism & Aging,
 368 Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria.
- 369 167. The New York Academy of Medicine, New York, NY, USA.
- 370 168. Alzheimer Scotland Dementia Research Centre, University of Edinburgh, 371 Edinburgh, UK.
- 372 169. Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, United Kingdom.
- 374 170. Population Health Research Institute, St George's, University of London, London, UK.
- 376 171. Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
- 172. Institute for Community Medicine, University Medicine Greifswald, Greifswald,Germany.
- 380 173. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands.

- 382 174. Dasman Diabetes Institute, Dasman, Kuwait.
- 383 175. Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland.
- 385 176. Department of Public Health, University of Helsinki, Helsinki, Finland.
- 386 177. Saudi Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia.
- 387 178. Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands.
- 388 179. Research Institute for Primordial Prevention of Non-communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran.
- 390 180. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany.
- 392 181. Department of Internal Medicine, University Hospital, CHUV, Lausanne, Switzerland.
- 394 182. Experimental Genetics Division, Sidra Medical and Research Center, Doha, Qatar.
- 395 183. Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Scotland, UK
- 397 184. Department of Biology, Faculty of Medicine, University of Split, Croatia.
- The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, UK.
- 400 186. Division of Cardiology, University Hospital, Basel, Switzerland.
- 401 187. Division of Cardiology, Department of Medicine, McMaster University, Hamilton, 402 Canada.
- 188. Institute of Genetic and Biomedical Research, National Research Council (CNR),
 404 Monserrato, Cagliari, Italy.
- 405 189. Department of Biomedical Sciences, University of Sassari, Sassari, Italy.
- 406 190. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.
- 408 191. Laboratory of Cardiovascular Science, NIA/NIH, Baltimore, MD, USA.
- 409 192. Department of Public Health and Primary Care, Leiden University Medical
 410 Center, Leiden, the Netherlands.
- 411 193. Labormedizinisches Zentrum Dr. Risch, Schaan, Liechtenstein.
- 412 194. Private University of the Principality of Liechtenstin, Triesen, Liechtenstein.
- 413 195. University Insitute of Clinical Chemistry, Inselspital, Bern University Hospital,
 414 University of Bern, Bern, Switzerland.
- 415 196. Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
- 417 197. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.
- 418 198. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, 419 USA.
- 420 199. Tennessee Valley Healthcare System (Nashville VA) & Vanderbilt University, TN, 421 USA.
- VA Boston Healthcare, Section of Cardiology and Department of Medicine,
 Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
- Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology
 and Health Services, University of Washington, Seattle, WA, USA.
- 426 202. Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA.
- 427 203. National Institute for Health Research Imperial Biomedical Research Centre,
- Imperial College Healthcare NHS Trust and Imperial College London, London, UK.

430 204. UK Dementia Research Institute (UK DRI) at Imperial College London, London, 431 UK 432 205. Health Data Research-UK London substantive site, London, U.K 433 434 435 Corresponding authors: Mark Caulfield (m.j.caulfield@qmul.ac.uk) and Paul Elliott 436 (p.elliott@imperial.ac.uk) 437 438 **Abstract** 439 High blood pressure is a highly heritable and modifiable risk factor for cardiovascular 440 disease. We report the largest genetic association study of blood pressure traits (systolic, 441 diastolic, pulse pressure) to date in over one million people of European ancestry. We 442 identify 535 novel blood pressure loci that not only offer new biological insights into blood 443 pressure regulation but also reveal shared genetic architecture between blood pressure and 444 lifestyle exposures. Our findings identify new biological pathways for blood pressure 445 regulation with potential for improved cardiovascular disease prevention in the future. 446

INTRODUCTION

- 448 High blood pressure (BP) is a leading heritable risk factor for stroke and coronary artery
- disease, responsible for an estimated 7.8 million deaths and 148 million disability life years
- lost worldwide in 2015 alone¹. Blood pressure is determined by complex interactions
- between life-course exposures and genetic background²⁻⁴. Previous genetic association
- studies have identified and validated variants at 274 loci with modest effects on population
- BP, explaining in aggregate ~3% of the trait variance⁵⁻¹².
- Here, we report genome-wide discovery analyses of BP traits systolic (SBP), diastolic (DBP)
- and pulse pressure (PP) in people of European ancestry drawn from UK Biobank (UKB)¹³
- 456 and the International Consortium of Blood Pressure-Genome Wide Association Studies
- 457 (ICBP)^{11,12}. We adopted a combination of a one- and two-stage study design to test common
- and low-frequency single nucleotide polymorphisms (SNPs) with minor allele frequency
- 459 (MAF) \geq 1% associated with BP traits (Fig. 1). In all, we studied over 1 million people of
- 460 European descent, including replication data from the US Million Veterans Program (MVP,
- 461 N=220,520)¹⁴ and the Estonian Genome Centre, University of Tartu (EGCUT, N=28,742)
- 462 Biobank¹⁵.
- 463 UKB is a prospective cohort study of ~500,000 richly phenotyped individuals, including BP
- measurements¹³, with genotyping by customized array and imputation from the Haplotype
- Reference Consortium (HRC) panel, yielding ~7 million SNPs (imputation quality score (INFO)
- \geq 0.1 and MAF \geq 1%)¹⁶. We performed genome-wide association studies (GWAS) of BP traits
- 467 (N=458,577 Europeans) under an additive genetic model¹⁷ (**Supplementary Table 1a**).
- Following LD-score regression¹⁸, genomic control (GC) was applied to the UKB data prior to
- 469 meta-analysis (Online methods).
- 470 In addition, we performed GWAS analyses for BP traits in newly extended ICBP GWAS data
- 471 comprising 77 independent studies for up to 299,024 Europeans genotyped with various
- arrays, and imputed to either the 1,000 Genomes Reference Panel or the HRC platforms
- 473 (Supplementary Table 1b). After QC we applied GC at the individual study level and
- 474 obtained summary effect sizes for ~7 million SNPs with INFO ≥ 0.3 and heterogeneity
- 475 Cochran's Q statistic¹⁹ filtered at $P \ge 1 \times 10^{-4}$ (Online Methods).
- 476 We then combined the UKB and ICBP GWAS results using inverse-variance weighted fixed
- 477 effects meta-analysis (Online Methods), giving a total discovery sample of up to 757,601
- 478 individuals²⁰.
- 479 In our two-stage design we attempted replication (in MVP and EGCUT, Supplementary
- **Table 1c**) of 1,062 SNPs at $P < 1 \times 10^{-6}$ from discovery with concordant effect direction
- between UKB and ICBP, using the sentinel SNP (i.e. SNP with smallest P-value at the locus)
- after excluding the HLA region (chr 6:25-34MB) and all SNPs in Linkage Disequilibrium (LD)
- 483 $(r^2 \ge 0.1)$ or ± 500 Kb from any previously validated BP-associated SNPs at the 274 published
- loci. Our replication criteria were genome-wide significance ($P < 5 \times 10^{-8}$) in the combined
- meta-analysis, P < 0.01 in the replication data and concordant direction of effect between
- 486 discovery and replication.

- 487 We additionally undertook a one-stage design to reduce type II error from the two-stage
- analysis. We used $P < 5 \times 10^{-9}$ as threshold from the discovery meta-analysis, i.e. an order of
- 489 magnitude more stringent than genome-wide significance²¹, and required an internal
- replication P < 0.01 in each of the UKB and ICBP GWAS analyses, with concordant direction
- 491 of effect, to minimize false positive findings.
- We carried out conditional analyses using genome-wide complex trait analysis (GCTA)²². We
- 493 then explored putative function of BP-associated signals using a range of in silico resources,
- 494 and evaluated co-occurrence of BP-associated loci with lifestyle exposures and other
- complex traits and diseases. Finally, we developed a genetic risk score (GRS) and assessed
- 496 impact of BP-associated variants on BP level, risk of hypertension (HTN), other
- 497 cardiovascular diseases and in other ethnicities.

RESULTS

498

519

- 499 We present a total of 535 novel loci (Fig.2, Supplementary Fig. 1): 325 loci claimed from the
- two-stage design (Supplementary Tables 2a-c) and an additional 210 claimed from our one-
- stage design with internal replication (Supplementary Tables 3a-c). Our two-stage design
- uniquely identified 121 variants, while 204 also met the one-stage criteria (Fig. 3a); large
- 503 numbers of loci would not have been detected by either the one- or two-stage designs
- alone (Fig. 3a). For SBP, the distributions of effect sizes are similar for the one-stage
- 505 (median = 0.219 mmHg per allele; Inter-Quartile Range (IQR) = 0.202-0.278) and two-stage
- loci (median = 0.224; IQR = 0.195-0.267) (P = 0.447) (Supplementary Fig. 2). Of the 210 loci
- found only in the one-stage analysis, 186 are also genome-wide significant ($P < 5 \times 10^{-8}$) in
- the combined meta-analysis, with all variants, except one, having concordant direction of
- effect between discovery and replication (Supplementary Tables 3a-c); of the remaining 24
- 510 SNPs, 10 still have concordant direction of effect.
- We find support in our data for all 274 previously published BP loci (Supplementary Fig. 1 &
- 512 **2** and **Supplementary Table 4**); >95% of the previously reported SNPs covered within our
- 513 data are genome-wide significant. Only 6 available SNPs did not reach Bonferroni-
- 514 significance, likely because they were originally identified in non-European ancestries (e.g.
- rs6749447, rs10474346, rs11564022), or from a gene-age interaction analysis (rs16833934).
- 516 In addition, we confirmed a further 92 previously reported, but not replicated, loci
- 517 (Supplementary Table 5)⁹; together with 274 previously reported loci confirmed, and 535
- 518 novel loci identified here, there are 901 BP-associated loci in total.

Novel genetic loci for blood pressure

- 520 Of the 535 independent novel loci, 363 SNPs were associated with one trait, 160 with two
- traits and 12 with all three BP traits (Fig. 3b). Using GCTA we additionally identified 163,
- 522 genome-wide significant, independent secondary signals with MAF ≥ 1% associated with BP
- (Supplementary Table 6), of which 19 SNPs are in LD ($r^2 \ge 0.1$) with previously reported
- secondary signals. This gives a total of 144 new secondary signals; hence we now report
- over 1,000 independent BP signals.
- The estimated SNP-wide heritability (h²) of BP traits in our data was 0.213, 0.212 and 0.194
- for SBP, DBP and PP respectively, with a gain in percentage of BP variance explained. For

- 528 example, for SBP, percentage variance explained increased from 2.8 % for the 274
- 529 previously published loci to 5.7% for SNPs identified at all 901 loci (Supplementary Table 7).

530 Functional analyses

- Our functional analyses approach is summarised in **Supplementary Figure 3**. First, for each
- of the 901 loci we annotated all SNPs (based on LD $r^2 \ge 0.8$) to the nearest gene within 5kb
- of a SNP, identifying 1333 genes for novel loci and 1272 genes for known loci. Then we
- 534 investigated these loci for tissue enrichment, DNase hypersensitivity site enrichment and
- pathway analyses. At 66 of the 535 novel loci we identified 97 non-synonymous SNPs,
- including 8 predicted to be damaging (Supplementary Table 8).
- We used chromatin interaction Hi-C data from endothelial cells (HUVEC)²³, neural
- progenitor cells (NPC), mesenchymal stem cells (HVMSC) and tissue from the aorta (HAEC)
- and adrenal gland²⁴ to identify distal associated genes. There were 498 novel loci that
- 540 contained a potential regulatory SNP and in 484 of these we identified long-range
- interactions in at least one of the tissues or cell types. We found several potential long-
- range target genes that do not overlap with the sentinel SNPs in the LD block. For example,
- the TGFB2 gene forms a 1.2Mb regulatory loop with SNPs in the SLC30A10 locus, and the
- 544 *TGFBR1* promoter forms a 100kb loop with the *COL15A1* locus **(Supplementary Table 8)**.
- Our eQTL analysis identified 60 novel loci with eQTLs in arterial and 20 in adrenal tissue
- 546 (Supplementary Table 9), substantially increasing those identified in our previously
- 547 published GWAS on ~140K UKB individuals¹⁰. An example is SNP rs31120122 which defines
- an aortic eQTL affecting expression of the *MED8* gene within the *SZT2* locus. In combination
- an additional and the expression of the *MLDB* gene within the 3212 locus. In combination
- with Hi-C interaction data in MSC, this supports a role for *MED8* in BP regulation, possibly
- 550 mediated through repression of smooth muscle cell differentiation. Hi-C interactions
- provide supportive evidence for involvement of a further 36 arterial eGenes (genes whose
- expression is affected by the eQTLs) that were distal to their eQTLs (e.g PPHLN1, ERAP2,
- 553 *FLRT2, ACVR2A, POU4F1*).
- 554 Using DeepSEA we found 198 SNPs in 121 novel loci with predicted effects on transcription
- factor binding or on chromatin marks in tissues relevant for BP biology, such as vascular
- tissue, smooth muscle and the kidney (Supplementary Table 8).
- 557 We used our genome-wide data at a false discovery rate (FDR) < 1% to robustly assess tissue
- enrichment of BP loci using DEPICT and identified enrichment across 50 tissues and cells
- (Supplementary Fig 4; Supplementary Table 10a). Enrichment was greatest for the
- cardiovascular system especially blood vessels ($P = 1.5 \times 10^{-11}$) and the heart ($P = 2.7 \times 10^{-5}$).
- Enrichment was high in adrenal tissue ($P = 3.7 \times 10^{-4}$) and, for the first time, we observed
- high enrichment in adipose tissues ($P = 9.8 \times 10^{-9}$) corroborated by eQTL enrichment
- analysis (P < 0.05) (Supplementary Fig. 4; Supplementary Table 10a). Evaluation of enriched
- mouse knockout phenotype terms also points to the importance of vascular morphology (*P*
- $= 6 \times 10^{-15}$) and development ($P = 2.1 \times 10^{-18}$) in BP. With addition of our novel BP loci, we
- 566 identified new findings from both the gene ontology and protein-protein interaction
- subnetwork enrichments, which highlight the TGF β ($P = 2.3 \times 10^{-13}$) and related SMAD
- pathways ($P = 7 \times 10^{-15}$) (Supplementary Table 10b, Supplementary Fig. 5b-d).

- We used FORGE²⁵ to investigate the regulatory regions for cell type specificity from DNase I
- 570 hypersensitivity sites, which showed strongest enrichment (P < 0.001) in the vasculature
- and highly vascularised tissues, as reported in previous BP genetic studies (Supplementary
- 572 **Fig. 6).**

592

608

Potential therapeutic targets

- 574 Ingenuity pathway analysis and upstream regulator assessment showed enrichment of
- 575 canonical pathways implicated in cardiovascular disease including pathways targeted by
- antihypertensive drugs (e.g. nitric oxide signalling) and also suggested some potential new
- 577 targets, such as relaxin signalling. Notably, upstream regulator analysis identified several BP
- 578 therapeutic targets such as angiotensinogen, calcium channels, progesterone, natriuretic
- 579 peptide receptor, angiotensin converting enzyme, angiotensin receptors and endothelin
- receptors (Supplementary Fig. 7).
- We developed a cumulative tally of functional evidence at each variant to assist in
- variant/gene prioritisation at each locus and present a summary of the vascular expressed
- 583 genes contained within the 535 novel loci, including a review of their potential druggability
- (Supplementary Fig. 8). The overlap between BP-associated genes and those associated
- 585 with antihypertensive drug targets further demonstrates new genetic support for known
- drug mechanisms. For example, we report five novel BP associations with targets of five
- antihypertensive drug classes (Supplementary Table 11), including the PKD2L1, SLC12A2,
- 588 CACNA1C, CACNB4 and CA7 loci targeted by potassium-sparing diuretics (amiloride), loop
- diuretics (bumetanide and furosemide), dihydropyridine, calcium channel blockers, non-
- 590 dihydropyridines and thiazide-like diuretics (chlortalidone) respectively. Notably in all but
- the last case, functional variants in these genes are the best candidates in each locus.

Concordance of BP variants and lifestyle exposures

- We examined association of sentinel SNPs at the 901 BP loci with BP-associated lifestyle
- 594 traits¹⁴ in UKB using either the Stanford Global Biobank Engine (N=327,302) or Gene ATLAS
- 595 (N=408,455). With corrected $P < 1 \times 10^{-6}$, we found genetic associations of BP variants with
- 596 daily fruit intake, urinary sodium and creatinine concentration, body mass index (BMI),
- weight, waist circumference, and intakes of water, caffeine and tea ($P = 1.03 \times 10^{-7}$ to P =
- 598 1.3 \times 10⁻⁴⁶). Specifically, SNP rs13107325 in *SLC39A8* is a novel locus for frequency of
- drinking alcohol ($P = 3.5 \times 10^{-15}$) and time spent watching TV ($P = 2.3 \times 10^{-11}$) as well as being
- associated with BMI ($P = 1.6 \times 10^{-33}$), weight ($P = 8.8 \times 10^{-16}$) and waist circumference ($P = 8.8 \times 10^{-16}$)
- 4.7×10^{-11}) (Supplementary Table 12). We used unsupervised hierarchical clustering for the
- 36 BP loci that showed at least one association at $P < 1 \times 10^{-6}$ with the lifestyle-related traits
- in UKB (Fig. 4). The heatmap summarises the locus-specific associations across traits and
- highlights heterogeneous effects with anthropometric traits across the loci examined. For
- example, it shows clusters of associations between BP-raising alleles and either increased or
- decreased adult height and weight. We note that some observed cross-trait associations are
- in counter-directions to those expected epidemiologically.

Association lookups with other traits and diseases

- We further evaluated cross-trait and disease associations using GWAS catalog²⁶, 609 PhenoScanner²⁷ and DisGeNET^{28,29}. The GWAS catalog and PhenoScanner search of 610 published GWAS showed that 77 of our 535 novel loci (using sentinel SNPs or proxies; r²≥ 611 612 0.8) are also significantly associated with other traits and diseases (Fig. 5, Supplementary 613 **Table 13**). We identified APOE as a highly cross-related BP locus showing associations with 614 lipid levels, cardiovascular-related outcomes and Alzheimer's disease, highlighting a 615 common link between cardiovascular risk and cognitive decline (Fig. 5). Other loci overlap with anthropometric traits, including BMI, birth weight and height (Fig 5) and with 616 617 DisGeNET terms related to lipid measurements, cardiovascular outcomes and obesity (Fig. 618 6).
- We did lookups of our sentinel SNPs in ¹H NMR lipidomics data on plasma (N=2,022) and 619 data from the Metabolon platform (N=1,941) in the Airwave Study³⁰, and used 620 PhenoScanner to test SNPs against published significant ($P < 5 \times 10^{-8}$) genome vs 621 622 metabolome-wide associations in plasma and urine (Online Methods). Ten BP SNPs show 623 association with lipid particle metabolites and a further 31 SNPs (8 also on PhenoScanner) 624 show association with metabolites on the Metabolon platform, highlighting lipid pathways, 625 amino acids (glycine, serine, glutamine), tri-carboxylic acid cycle intermediates 626 (succinylcarnitine) and drug metabolites (Supplementary Tables 14 and 15). These findings 627 suggest a close metabolic coupling of BP regulation with lipid and energy metabolism.

Genetic risk of increased blood pressure, hypertension and cardiovascular disease

- 629 A weighted GRS for BP levels across all 901 loci was associated with a 10.4 mmHg higher, sex-adjusted mean SBP in UK Biobank comparing the upper and lower quintiles of the GRS 630 distribution (95% CI: 10.2 to 10.6 mm Hg, $P < 1 \times 10^{-300}$) and with 12.9 mmHg difference in 631 SBP (95% CI: 12.6 to 13.1, $P < 1 \times 10^{-300}$) comparing the upper and lower deciles (**Fig. 7a**, 632 Supplementary Table 16). In addition, we observed over two-fold sex-adjusted higher risk 633 of hypertension (OR 2.66; 95% CI: 2.60 to 2.72; $P < 1 \times 10^{-300}$) between the upper and lower 634 635 quintiles of the GRS in UK Biobank (Fig. 7). Sensitivity analyses in the independent Airwave 636 cohort gave similar results (Supplementary Table 17).
- We also show that the GRS is associated with increased, sex-adjusted risk of incident stroke, myocardial infarction and all incident cardiovascular outcomes, comparing upper and lower deciles of the GRS distribution, with odds ratios of 1.47 (95% CI: 1.35 to 1.59, $P = 1.12 \times 10^{-6}$) 1.50 (95% CI: 1.28 to 1.76, $P = 7.99 \times 10^{-7}$) and 1.52 (95% CI: 1.26 to 1.82, $P = 7.4 \times 10^{-6}$) respectively (**Fig. 7b, Supplementary Table 16**).

Extending analyses to other ancestries

628

642

We examined associations with BP of both individual SNPs and the GRS among unrelated individuals of African and South Asian descent in UKB, for the 901 known and novel loci. Compared to Europeans, 62.4%, 62.5% and 64.8% of the variants among Africans (N=7,782), and 74.2%, 72.3% and 75% South Asians (N=10,323) have concordant direction of effect for SBP, DBP and PP respectively (Supplementary Table 18; Supplementary Fig. 9). Pearson correlation coefficients with effect estimates in Europeans were r²= 0.37 and 0.78 for Africans and South Asians respectively (Supplementary Fig. 10). We then applied the

- 650 European-derived GRS findings to unrelated Africans (N=6,970) and South Asians (N=8,827).
- BP variants in combination were associated with 6.1 mmHg (95% CI: 4.50 to 7.65; $P = 4.9 \times 10^{-2}$
- 652 10^{-14}) and 7.4 mmHg (95% CI: 6.00 to 8.69; $P = 1.7 \times 10^{-26}$) higher, sex-adjusted mean systolic
- 653 pressure among Africans and South Asians, respectively, comparing upper and lower
- quintiles of the GRS distribution (Supplementary Tables 19a and 19b).

DISCUSSION

- 656 Our study of over 1 million people offers an important step forward in understanding the 657 genetic architecture of BP. We identified over 1,000 independent signals at 901 loci for BP 658 traits, and the 535 novel loci more than triples the number of BP loci and doubles the 659 percentage variance explained, illustrating the benefits of large-scale biobanks. By explaining 27% of the estimated heritability for BP, we make major inroads into the missing 660 heritability influencing BP level in the population³¹. The novel loci open the vista of entirely 661 662 new biology and highlight gene regions in systems not previously implicated in BP 663 regulation. This is particularly timely as global prevalence of people with SBP over 110-115 664 mm Hg, above which cardiovascular risk increases in a continuous graded manner, now exceeds 3.5 billion, of whom over 1 billion are within the treatment range ^{32,33}. 665
- 666 Our functional analysis highlights the role of the vasculature and associated pathways in the 667 genetics underpinning BP traits. We show a role for several loci in the transforming growth 668 factor beta (TGF β) pathway including SMAD family genes and the TGF β gene locus itself. This pathway affects sodium handling in the kidney, ventricular remodelling, while plasma 669 levels of TGFB have recently been correlated with hypertension (Fig. 8)^{34,35}. The activin A 670 receptor type 1C (ACVR1C) gene mediates the effects of the TGFB family of signalling 671 672 molecules. A BP locus contains the Bone Morphogenetic Protein 2 (BMP2) gene in the TGFβ 673 pathway, which prevents growth suppression in pulmonary arterial smooth muscle cells and is associated with pulmonary hypertension³⁶. Another BP locus includes the Kruppel-like 674 family 14 (KLF14) gene of transcription factors, induced by low levels of TGFβ receptor II 675 676 gene expression, and which has also been associated with type 2 diabetes, hypercholesterolaemia and atherosclerosis³⁷. 677
- 678 Our analysis shows enrichment of BP gene expression in the adrenal tissue. Autonomous aldosterone production by the adrenal glands is thought to be responsible for 5-10% of all 679 hypertension, rising to ~20% amongst people with resistant hypertension³⁸. Some of our 680 novel loci are linked functionally to aldosterone secretion^{39,40}. For example, the CTNNB1 681 locus encodes β-catenin, the central molecule in the canonical Wnt signalling system, 682 required for normal adrenocortical development^{41,42}. Somatic adrenal mutations of this 683 gene that prevent serine/threonine phosphorylation lead to hypertension through 684 generation of aldosterone-producing adenomas^{43,44}. 685
- Our novel loci also include genes involved in vascular remodelling, such as vascular endothelial growth factor A (*VEGFA*), the gene product of which induces proliferation, migration of vascular endothelial cells and stimulates angiogenesis. Disruption of this gene in mice resulted in abnormal embryonic blood vessel formation, while allelic variants of this gene have been associated with microvascular complications of diabetes, atherosclerosis and the antihypertensive response to enalapril⁴⁵. We previously reported a fibroblast

692 growth factor (*FGF5*) gene locus in association with BP¹⁰. Here, we additionally identify a 693 new BP locus encoding FGF9, which is linked to enhanced angiogenesis and vascular smooth 694 muscle cell differentiation by regulating *VEGFA* expression.

695 Several of our novel loci contain lipid-related genes consistent with the observed strong 696 associations among multiple cardio-metabolic traits. For example, the apolipoprotein E 697 gene (APOE) encodes the major apoprotein of the chylomicron. Recently, APOE serum levels have been correlated with SBP in population-based studies and in murine knockout models; 698 disruption of this gene led to atherosclerosis and hypertension^{46,47}. A second novel BP locus 699 contains the low-density lipoprotein receptor-related protein 4 (LRP4) gene which may be a 700 701 target for APOE and is strongly expressed in the heart in mice and humans. In addition, we 702 identified a novel locus including the apolipoprotein L domain containing 1 gene (APOLD1) 703 that is highly expressed in the endothelium of developing tissues (particularly heart) during 704 angiogenesis.

705

706

707

708

709

710

711712

713

714

715716

717

718

Many of our novel BP loci encode proteins which may modulate vascular tone or signalling. For example, the locus containing urotensin-2 receptor (*UTS2R*) gene encodes a class A rhodopsin family G-protein coupled-receptor that upon activation by the neuropeptide urotensin II, produces profound vasoconstriction. One novel locus for SBP contains the relaxin gene, encoding a G-protein coupled receptor, with roles in vasorelaxation and cardiac function; it signals by phosphatidylinositol 3-kinase (PI3K)^{48,49}, an enzyme which inhibits vascular smooth muscle cell proliferation and neo-intimal formation⁵⁰. We identify the *PI3K* gene here as a novel BP locus. We also identify the novel *RAMP2* locus which encodes an adrenomedullin receptor⁵¹; we previously identified the adrenomedullin (*ADM*) gene as a BP locus¹². Adrenomedullin is known to exert differential effects on BP in the brain (vasopressor) and the vasculature (vasodilator). In addition, a locus containing Rho guanine nucleotide exchange factor 25 (*ARHGEF25*) gene generates a factor that interacts with Rho GTPases involved in contraction of vascular smooth muscle and regulation of responses to angiotensin II⁵².

719 We evaluated the 901 BP loci for extant or potentially druggable targets. Loci encoding 720 MARK3, PDGFC, TRHR, ADORA1, GABRA2, VEGFA and PDE3A are within systems with 721 existing drugs not currently linked to a known antihypertensive mechanism; they may offer 722 repurposing opportunities e.g. detection of SLC5A1 as the strongest repurposing candidate 723 in a new BP locus targeted by the type-2 diabetes drug canagliflozin. This is important as 724 between 8-12% of patients with hypertension exhibit resistance or intolerance to current 725 therapies and repositioning of a therapy with a known safety profile may reduce 726 development costs.

This study strengthens our previously reported GRS analysis indicating that all BP elevating alleles combined could increase systolic BP by 10 mm Hg or more across quintiles or deciles of the population distribution, substantially increasing risk of cardiovascular events¹⁰. We previously suggested that genotyping BP elevating variants in the young may lead to targeted lifestyle intervention in early life that might attenuate the BP rise at older ages¹⁰.

- 732 We identified several BP-associated loci that are also associated with lifestyle traits,
- suggesting shared genetic architecture between BP and lifestyle exposures ⁵³. We adjusted
- our BP GWAS analyses for BMI to control for possible confounding effects, though we
- acknowledge the potential for collider bias⁵⁴. Nonetheless, our findings of possible genetic
- overlap between loci associated with BP and lifestyle exposures could support renewed
- 737 focus on altering specific lifestyle measures known to affect BP⁵⁵.
- 738 Despite smaller sample sizes, we observed high concordance with direction of effects on BP
- traits of BP variants in Africans (> 62%) and South Asians (> 72%). The GRS analyses show
- that, in combination, BP variants identified in European analyses are associated with BP in
- non-European ancestries, though effect sizes were 30-40% smaller.
- Our use of a two- and one-stage GWAS design illustrates the value of this approach to
- 743 minimize the effects of stochastic variation and heterogeneity. The one-stage approach
- included signals that had independent and concordant support (P < 0.01) from both UKB
- and ICBP, reducing the impact of winners' curse on our findings. Indeed, all but two of the
- 746 210 SNPs discovered in the one-stage analysis reach $P < 5 \times 10^{-6}$ in either UKB or ICBP. To
- 747 further minimize the risk of reporting false positive loci within our one-stage design, we set
- 748 a stringent overall discovery meta-analysis *P*-value threshold of $P < 5 \times 10^{-9}$, an order of
- 749 magnitude smaller than a genome-wide significance P-value, in line with thresholds
- recommended for whole genome sequencing²². We found high concordance in direction of
- 751 effects between discovery data in the one-stage approach and the replication resources,
- 752 with similar distributions of effect sizes for the two approaches. We note that 24 of the
- one-stage SNPs which reached $P < 5 \times 10^{-9}$ in discovery failed to reach genome-wide
- 754 significance ($P < 5 \times 10^{-8}$) in the combined meta-analysis of discovery and replication
- resources, and hence may still require further validation in future, larger studies.
- 756 The new discoveries reported here more than triple the number of loci for BP to a total of
- 757 901 and represent a substantial advance in understanding the genetic architecture of BP.
- 758 The identification of many novel genes across the genome, could partly support an
- 759 omnigenic model for complex traits where genome-wide association of multiple
- 760 interconnected pathways is observed. However, our strong tissue enrichment shows
- particular relevance to the biology of BP and cardiovascular disease⁵⁶, suggesting trait-
- specificity, which could argue against an omnigenic model. Our confirmation of the impact
- of these variants on BP level and cardiovascular events, coupled with identification of
- shared risk variants for BP and adverse lifestyle could contribute to an early life precision
- medicine strategy for cardiovascular disease prevention.
- 766 **URLs**
- 767 FORGE: http://browser.1000genomes.org/Homo_sapiens/UserData/Forge?db=core
- 768 Fantom5 data: http://fantom.gsc.riken.jp/5/
- 769 ENCODE DNase I data: (wgEncodeAwgDnaseMasterSites; accessed using Table browser)
- 770 ENCODE cell type data: http://genome.ucsc.edu/ENCODE/cellTypes.html.
- 771 GTEx: www.gtexportal.org
- 772 DeepSEA: http://deepsea.princeton.edu/
- 773 WebGetstalt: http://www.webgestalt.org

- 774 IPA: www.qiagen.com/ingenuity
- 775 Mouse Genome Informatics (MGI): http://www.informatics.jax.org/batch
- 776 Drug Gene Interaction database: www.dgidb.org
- 777 PhenoScanner: http://www.phenoscanner.medschl.cam.ac.uk (Phenoscanner integrates
- results from the GWAS catalogue: https://www.ebi.ac.uk/gwas/ and GRASP:
- 779 https://grasp.nhlbi.nih.gov/)
- 780 DisGeNEt: http://www.disgenet.org
- 781 GeneATLAS: http//geneatlas.roslin.ed.ac.uk
- 782 Global Biobank Engine: https://biobankengine.stanford.edu

Acknowledgements

784 785

820

786 H.R.W. was funded by the National Institute for Health Research (NIHR) as part of the 787 portfolio of translational research of the NIHR Biomedical Research Centre at Barts and The 788 London School of Medicine and Dentistry, D.M-A is supported by the Medical Research 789 Council [grant number MR/L01632X.1]. B.M. holds an MRC eMedLab Medical Bioinformatics 790 Career Development Fellowship, funded from award MR/L016311/1. H.G. was funded by 791 the NIHR Imperial College Health Care NHS Trust and Imperial College London Biomedical 792 Research Centre. C.P.C. was funded by the National Institute for Health Research (NIHR) as 793 part of the portfolio of translational research of the NIHR Biomedical Research Center at 794 Barts and The London School of Medicine and Dentistry, S.T. was supported by Canadian 795 Institutes of Health Research; Université Laval (Quebec City, Canada). G.P. was supported by 796 Canada Research Chair in Genetic and Molecular Epidemiology and CISCO Professorship in 797 Integrated Health Biosystems. I.K. was supported by the EU PhenoMeNal project (Horizon 798 2020, 654241). C.P.K. is supported by grant U01DK102163 from the NIH-NIDDK, and by 799 resources from the Memphis VA Medical Center. C.P.K. is an employee of the US 800 Department of Veterans affairs. Opinions expressed in this paper are those of the authors' 801 and do not necessarily represent the opinion of the Department of Veterans Affairs. S.D. 802 was supported for this work by grants from the European Research Council (ERC), the EU 803 Joint Programme - Neurodegenerative Disease Research (JPND), the Agence Nationale de la 804 Recherche (ANR). T.B., J.MART., V.V., A.F.W. and C.H. were supported by a core MRC grant 805 to the MRCHGU QTL in Health and Disease research programme. M.BOE is supported by NIH 806 grant R01-DK062370. H.W. and A.G. acknowledge support of the Tripartite 807 Immunometabolism Consortium [TrIC], Novo Nordisk Foundation (grant NNF15CC0018486). 808 N.V. was supported by Marie Sklodowska-Curie GF grant (661395) and ICIN-NHI. C.M. is 809 funded by the MRC AimHy (MR/M016560/1) project grant. M.A.N participation is supported 810 by a consulting contract between Data Tecnica International and the National Institute on 811 Aging, NIH, Bethesda, MD, USA. M.BR., M.CO., I.G., P.G., G.G, A.MO., A.R., D.V., C.M.B., 812 C.F.S., M.T., D.T. were supported by the Italian Ministry of Health RF2010 to Paolo 813 Gasparini, RC2008 to Paolo Gasparini. D.I.B. is supported by the Royal Netherlands 814 Academy of Science Professor Award (PAH/6635), J.C.C. is supported by the Singapore 815 Ministry of Health's National Medical Research Council under its Singapore Translational 816 Research Investigator (STaR) Award (NMRC/STaR/0028/2017). C.C., P.B.M and M.R.B were 817 funded by the National Institutes for Health Research (NIHR) as part of the portfolio of 818 translational research of the NIHR Biomedical Research Centre at Barts. T.F. is supported by 819 the NIHR Biomedical Research Centre, Oxford. M.R. is recipient from China Scholarship

Council (No. 2011632047). C.L. was supported by the Medical Research Council UK

```
821
       (G1000143; MC UU 12015/1; MC PC 13048; MC U106179471), Cancer Research UK
822
       (C864/A14136), EU FP6 programme (LSHM CT 2006 037197). G.B.E is supported by the
823
       Swiss National Foundation SPUM project FN 33CM30-124087, Geneva University, and the
824
       Fondation pour Recherches Médicales, Genève. C.M.L is supported by the Li Ka Shing
825
       Foundation, WT-SSI/John Fell funds and by the NIHR Biomedical Research Centre, Oxford,
826
       by Widenlife and NIH (CRR00070 CR00.01). R.J.F.L. is supported by the NIH (R01DK110113,
827
       U01HG007417, R01DK101855, R01DK107786). D.O.M-K. is supported by Dutch Science
828
       Organization (ZonMW-VENI Grant 916.14.023). M.M was supported by the National
829
       Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical
830
       Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College
831
       London. H.W. and M.F. acknowledge the support of the Wellcome Trust core award
832
       (090532/Z/09/Z) and the BHF Centre of Research Excellence (RE/13/1/30181). A.G, H.W.
833
       acknowledge European Union Seventh Framework Programme FP7/2007-2013 under grant
834
       agreement no. HEALTH-F2-2013-601456 (CVGenes@Target) & and A.G., the Wellcome Trust
835
       Institutional strategic support fund. L.R. was supported by Forschungs- und Förder-Stiftung
836
       INOVA, Vaduz, Liechtenstein. M.TO. is supported by British Heart Foundation
837
       (PG/17/35/33001). P.S. is recipient of an NIHR Senior Investigator Award and is supported
838
       by the Biomedical Research Centre Award to Imperial College Healthcare NHS Trust. P.v.d.H.
839
       was supported by ICIN-NHI and Marie Sklodowska-Curie GF (call: H2020-MSCA-IF-2014,
840
       Project ID: 661395). N.J.W. was supported by the Medical Research Council UK (G1000143;
841
       MC UU 12015/1; MC PC 13048; MC U106179471), Cancer Research UK (C864/A14136),
842
       EU FP6 programme (LSHM CT 2006 037197). E.Z. was supported by the Wellcome Trust
843
       (WT098051). J.N.H. was supported by the Vanderbilt Molecular and Genetic Epidemiology
844
       of Cancer (MAGEC) training program, funded by T32CA160056 (PI: X.-O. Shu) and by VA
845
       grant 1I01CX000982. A.G. was supported by VA grant 1I01CX000982. T.L.E. and D.R.V.E.
846
       were supported by grant R21HL121429 from NIH/NHLBI. A.M.H. was supported by VA
847
       Award #I01BX003360. C.J.O. was supported by the VA Boston Healthcare, Section of
848
       Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical
849
       School. The MRC/BHF Cardiovascular Epidemiology Unit is supported by the UK Medical
850
       Research Council [MR/L003120/1]; British Heart Foundation [RG/13/13/30194]; and UK
851
       National Institute for Health Research Cambridge Biomedical Research Centre. J.DA is a
852
       British Heart Foundation Professor and NIHR Senior Investigator, L.V.W. holds a
853
       GlaxoSmithKline/British Lung Foundation Chair in Respiratory Research. P.E. acknowledges
854
       support from the NIHR Biomedical Research Centre at Imperial College Healthcare NHS
855
       Trust and Imperial College London, the NIHR Health Protection Research Unit in Health
856
       Impact of Environmental Hazards (HPRU-2012-10141), and the Medical Research Council
857
       (MRC) and Public Health England (PHE) Centre for Environment and Health
858
       (MR/L01341X/1). P.E. is a UK Dementia Research Institute (DRI) professor, UK DRI at
859
       Imperial College London, funded by the MRC, Alzheimer's Society and Alzheimer's Research
860
       UK. He is also associate director of Health Data Research-UK London funded by a
861
       consortium led by the Medical Research Council. M.J.C. was funded by the National Institute
862
       for Health Research (NIHR) as part of the portfolio of translational research of the NIHR
863
       Biomedical Research Center at Barts and The London School of Medicine and Dentistry.
864
       M.J.C. is a National Institute for Health Research (NIHR) senior investigator and this work is
865
       funded by the MRC eMedLab award to M.J.C. and M.R.B. and the NIHR Biomedical Research
866
       Centre at Barts.
```

- 867 This research has been conducted using the UK Biobank Resource under Application
- Numbers 236 and 10035. This research was supported by the British Heart Foundation
- 869 (grant SP/13/2/30111). Large-scale comprehensive genotyping of UK Biobank for
- 870 cardiometabolic traits and diseases: UK CardioMetabolic Consortium (UKCMC).
- 871 Computing: This work was enabled using the computing resources of the i) UK MEDical
- 872 BIOinformatics partnership aggregation, integration, visualisation and analysis of large,
- 873 complex data (UK MED-BIO) which is supported by the Medical Research Council [grant
- number MR/L01632X/1] and ii) the MRC eMedLab Medical Bioinformatics Infrastructure,
- supported by the Medical Research Council [grant number MR/L016311/1].

876 Author contributions

- 878 *Central analysis:* E.E., H.R.W., D.M-A., B.M., R.P., H.G., G.N., N.D., C.P.C., I.K., F.N., M.E.,
- 879 K.W., E.T. L.V.W.
- 880 Writing of the manuscript: E.E., H.R.W., D.M-A., B.M., R.P., H.G., I.T., M.R.B., L.V.W., P.E.,
- 881 M.J.C. (with group leads EE, H.R.W, L.V.W., P.E., M.J.C.)
- 882 ICBP-Discovery contributor: (3C-Dijon) S.D., M.S., P.AM., G.C., C.T.; (AGES-Reykjavik) V.GU.,
- 883 L.J.L., A.V.S., T.B.H.; (ARIC) D.E.A., E.B., A.CH. A.C.M., P.N.; (ASCOT) N.R.P., D.C.S., A.S.,
- 884 S.THO., P.B.M., P.S., M.J.C., H.R.W.; (ASPS) E.H., Y.S., R.S., H.S.; (B58C) D.P.S., BHSA.J.,
- N.SHR.; (BioMe (formerly IPM)) E.P.B., Y.LU., R.J.F.L.; (BRIGHT) J.C., M.F., M.J.B., P.B.M.,
- 886 M.J.C., H.R.W.; (CHS) J.C.B., K.R., K.D.T., B.M.P.; (Cilento study) M.C., T.NU., D.R., R.SO.;
- (COLAUS) M.B., Z.K., P.V.; (CROATIA Korcula) J.MART., A.F.W.; (CROATIA SPLIT) I.KO., O.P.,
- 888 T.Z.; (CROATIA Vis) J.E.H., I.R., V.V.; (EPIC) K-T.K., R.J.F.L., N.J.W.; (EPIC-CVD) W-Y.L., P.SU.,
- 889 A.S.B., J.DA., J.M.M.H.; (EPIC-Norfolk, Fenland-OMICS, Fenland-GWAS) J-H.Z.; (EPIC-Norfolk,
- 890 Fenland-OMICS, Fenland-GWAS, InterAct-GWAS) J.L., C.L., R.A.S., N.J.W.; (ERF) N.A., B.A.O.,
- 891 C.M.v.D.; (Fenland-Exome, EPIC-Norfolk-Exome) S.M.W., FHSS-J.H., D.L.; (FINRISK
- 892 (COROGENE CTRL)) P.J., K.K., M.P., A-P.S.; (FINRISK PREDICT CVD) A.S.H., A.P., S.R., V.S.;
- 893 (FUSION) A.U.J, M.BOE., F.C., J.T., (GAPP) S.T., G.P., D.CO., L.R.; (Generation Scotland
- 894 (GS:SFHS)) T.B., C.H., A.C., S.P.; (GODARTS) N.S., A.S.F.D., A.D.M., C.N.A.P.; (GRAPHIC) P.S.B.,
- 895 C.P.N., N.J.SA., M.D.T.; (H2000 CTRL) A.JU., P.K., S.KO., T.N.; (HABC) Y.L., M.A.N., T.B.H.;
- 896 (HCS) J.R.A., E.G.H., C.O., R.J.SC.; (HTO) K.L.A., H.J.C., B.D.K., M.TO, C.MA.; (ICBP-SC) G.A.,
- 897 T.F., M-R.J., A.D.J., M.LA., C.N.; (INGI-CARL) I.G., G.G., A.MO., A.R.; (INGI-FVG) M.BR., M.CO.,
- 898 P.G., D.V.; (INGI-VB) C.M.B., C.F.S., D.T., M.T.; (JUPITER) F.G., L.M.R., P.M.R., D.I.C.; (KORA
- 899 S3) C.G., M.L., E.O., S.S.; (KORA S4) A.PE., J.S.R.; (LBC1921) S.E.H., D.C.M.L., A.PA., J.M.S.;
- 900 (LBC1936) G.D., I.J.D., A.J.G., L.M.L.; (Lifelines) N.V., M.H.d.B., M.A.S., P.v.d.H.; (LOLIPOP)
- 901 J.C.C., J.S.K., B.L., W.Z.; (MDC) P.A., O.M.; (MESA) X.G., W.P., J.I.R., J.Y.; (METSIM) A.U.J.,
- 902 M.LAA.; (MICROS) F.D.G.M., A.A.H., P.P.P.; (MIGEN) R.E., S.K., J.M., D.SI.; (NEO) R.L., R.d.M.,
- 903 R.N., D.O.M-K.; (NESDA) Y.M., I.M.N., B.W.J.H.P., H.SN.; (NSPHS) S.E., U.G., Å.JO.; (NTR)
- 904 D.I.B., E.J.d.G., J-J.H., G.W.; (ORCADES) H.C., P.K.J., S.H.W., J.F.W.; (PIVUS) L.LI., C.M.L., J.S.,
- 905 A.M.; (Prevend) N.V., P.v.d.H.; (PROCARDIS) M.F., A.G., H.W.; (PROSPER) J.DE., J.W.J., D.J.S.,
- 906 S.TR.; (RS) O.H.F., A.HO., A.U., G.C.V.; (SardiNIA) J.D., Y.Q., F.CU., E.G.L.; (SHIP) M.D., R.R.,
- 907 A.T., U.V.; (STR) M.FR., A.H., R.J.S., E.I.; (TRAILS) C.A.H., A.J.O., H.R., P.J.v.d.M.; (TwinsUK)

- 908 M.M., C.M., T.D.S.; (UKHLS) B.P.P., E.Z.; (ULSAM) V.G., A.P.M., A.M., E.I.; (WGHS) F.G.,
- 909 L.M.R., P.M.R., D.I.C.; (YFS) M.K., T.L., L-P.L., O.T.R.
- 910 Replication study contributor: (MVP) J.N.H., A.G., D.R.V.E., Y.V.S., K.C., J.M.G., P.W.F.W.,
- 911 P.S.T., C.P.K., A.M.H., C.J.O., T.L.E.; (EGCUT) T.E., R.M., L.M. A.ME.
- 912 **Airwave Health Monitoring Study:** E.E, H.G, A-C.V., R.P., I.K., I.T., P.E.
- 913 All authors critically reviewed and approved the final version of the manuscript
- 914 Conflicts/Disclosures

- 916 K.W. is a Commercial partnerships manager for Genomics England, a UK Government
- 917 Company
- 918 M.A.N. consults for Illumina Inc, the Michael J. Fox Foundation and University of California
- 919 Healthcare among others.
- 920 A.S.B. has received grants outside of this work from Merck, Pfizer, Novartis, AstraZeneca,
- 921 Biogen and Bioverativ and personal fees from Novartis
- 922 J.DA. has the following competing interests: Pfizer Population Research Advisory Panel
- 923 (grant), AstraZeneca (grant), Wellcome Trust (grant), UK Medical Research Council (grant),
- 924 Pfizer(grant), Novartis (grant), NHS Blood and Transplant(grant), National Institute of Health
- 925 Research (grant), UK MEDICAL RESEARCH COUNCIL (grant), BRITISH HEART
- 926 FOUNDATION(grant), UK NATIONAL INSTITUTE OF HEALTH RESEARCH (grant), EUROPEAN
- 927 COMMISSION (grant), Merck Sharp and Dohme UK Atherosclerosis (personal fees), Novartis
- 928 Cardiovascular and Metabolic Advisory Board (personal fees), British Heart Foundation
- 929 (grant), European Research Council (grant), Merck (grant).
- 930 B.M.P. serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering
- Ommittee of the Yale Open Data Access Project funded by Johnson & Johnson.
- 932 M.J.C. is Chief Scientist for Genomics England, a UK Government company.

933

- The views expressed in this manuscript are those of the authors and do not necessarily
- 935 represent the views of the National Heart, Lung, and Blood Institute; the National Institutes
- of Health; or the U. S. Department of Health and Human Services. This publication does not
- 937 represent the views of the Department of Veterans Affairs or the United States Government.

938

References

940

- 941 1. Forouzanfar, M.H. *et al.* Global Burden of Hypertension and Systolic Blood Pressure 942 of at Least 110 to 115 mm Hg, 1990-2015. *JAMA* **317**, 165-182 (2017).
- 943 2. Munoz, M. *et al.* Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet* **48**, 980-3 (2016).
- 945 3. Poulter, N.R., Prabhakaran, D. & Caulfield, M. Hypertension. *Lancet* **386**, 801-12 (2015).

- 947 4. Feinleib, M. *et al.* The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. *Am J Epidemiol* **106**, 284-5 (1977).
- Cabrera, C.P. *et al.* Exploring hypertension genome-wide association studies findings
 and impact on pathophysiology, pathways, and pharmacogenetics. *Wiley Interdiscip Rev Syst Biol Med* 7, 73-90 (2015).
- 952 6. Ehret, G.B. *et al.* The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet* **48**, 1171-1184 (2016).
- 954 7. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-1161 (2016).
- 956 8. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood 957 pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
- 958 9. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* **49**, 54-64 (2017).
- 961 10. Warren, H.R. *et al.* Genome-wide association analysis identifies novel blood pressure 962 loci and offers biological insights into cardiovascular risk. *Nat Genet* **49**, 403-415 963 (2017).
- 964 11. Wain, L.V. et al. Novel Blood Pressure Locus and Gene Discovery Using Genome 965 Wide Association Study and Expression Data Sets From Blood and the Kidney.
 966 Hypertension (2017).
- 967 12. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.*968 Genetic variants in novel pathways influence blood pressure and cardiovascular
 969 disease risk. *Nature* **478**, 103-9 (2011).
- 970 13. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
- 973 14. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).
- 975 15. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* **44**, 1137-47 (2015).
- 977 16. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. 978 *Nat Genet* **48**, 1279-83 (2016).
- Loh, P.R. et al. Efficient Bayesian mixed-model analysis increases association power
 in large cohorts. Nat Genet 47, 284-90 (2015).
- 981 18. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
- 983 19. Ioannidis, J.P., Patsopoulos, N.A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* **2**, e841 (2007).
- 985 20. Evangelou, E. & Ioannidis, J.P. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* **14**, 379-89 (2013).
- 987 21. Pulit, S.L., de With, S.A. & de Bakker, P.I. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genet Epidemiol* **41**, 145-151 (2017).
- 990 22. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
- 992 23. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).

- 994 24. Schmitt, A.D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* **17**, 2042-2059 (2016).
- 996 25. Dunham, I.K., E.; Iotchkova, V.; Morganella, S.; Birney, E. FORGE: A tool to discover cell specific enrichments of GWAS associated SNPs in regulatory regions.

 998 F1000Research 4(2015).
- 999 26. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
- 1001 27. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207-3209 (2016).
- Pinero, J. *et al.* DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* **2015**, bav028 (2015).
- 1005 29. Pinero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 1007 (2017).
- 1008 30. Elliott, P. *et al.* The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. *Environ Res* **134**, 280-5 (2014).
- 1010 31. Ehret, G.B. & Caulfield, M.J. Genes for blood pressure: an opportunity to understand hypertension. *Eur Heart J* **34**, 951-61 (2013).
- 1012 32. Blood Pressure Lowering Treatment Trialists, C. *et al.* Blood pressure-lowering treatment based on cardiovascular risk: a meta-analysis of individual patient data. *Lancet* **384**, 591-8 (2014).
- 33. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk
 1016 assessment of 79 behavioural, environmental and occupational, and metabolic risks
 1017 or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease
 1018 Study 2015. Lancet 388, 1659-1724 (2016).
- Nakao, E. *et al.* Elevated Plasma Transforming Growth Factor beta1 Levels Predict
 the Development of Hypertension in Normotensives: The 14-Year Follow-Up Study.
 Am J Hypertens 30, 808-814 (2017).
- 1022 35. Feng, W., Dell'Italia, L.J. & Sanders, P.W. Novel Paradigms of Salt and Hypertension. *J Am Soc Nephrol* **28**, 1362-1369 (2017).
- 1024 36. International PPH Consortium *et al.* Heterozygous germline mutations in BMPR2,
 1025 encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. *Nat* 1026 *Genet* 26, 81-4 (2000).
- 1027 37. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-1028 scale association analysis. *Nat Genet* **42**, 579-89 (2010).
- Douma, S. *et al.* Prevalence of primary hyperaldosteronism in resistant hypertension: a retrospective observational study. *Lancet* **371**, 1921-6 (2008).
- 1031 39. Rossi, G.P. *et al.* A prospective study of the prevalence of primary aldosteronism in 1,125 hypertensive patients. *J Am Coll Cardiol* **48**, 2293-300 (2006).
- Calhoun, D.A., Nishizaka, M.K., Zaman, M.A., Thakkar, R.B. & Weissmann, P.
 Hyperaldosteronism among black and white subjects with resistant hypertension.
- 1035 *Hypertension* **40**, 892-6 (2002).
- 1036 41. Drelon, C., Berthon, A., Mathieu, M., Martinez, A. & Val, P. Adrenal cortex tissue 1037 homeostasis and zonation: A WNT perspective. *Mol Cell Endocrinol* **408**, 156-64 1038 (2015).
- 1039 42. El Wakil, A. & Lalli, E. The Wnt/beta-catenin pathway in adrenocortical development and cancer. *Mol Cell Endocrinol* **332**, 32-7 (2011).

- 1041 43. Teo, A.E. *et al.* Pregnancy, Primary Aldosteronism, and Adrenal CTNNB1 Mutations. 1042 *N Engl J Med* **373**, 1429-36 (2015).
- 1043 44. Tissier, F. et al. Mutations of beta-catenin in adrenocortical tumors: activation of the
 1044 Wnt signaling pathway is a frequent event in both benign and malignant
 1045 adrenocortical tumors. Cancer Res 65, 7622-7 (2005).
- 1046 45. Oliveira-Paula, G.H. *et al.* Polymorphisms in VEGFA gene affect the antihypertensive responses to enalapril. *Eur J Clin Pharmacol* **71**, 949-57 (2015).
- 1048 46. Yang, R. *et al.* Hypertension and endothelial dysfunction in apolipoprotein E knockout mice. *Arterioscler Thromb Vasc Biol* **19**, 2762-8 (1999).
- Sofat, R. et al. Circulating Apolipoprotein E Concentration and Cardiovascular
 Disease Risk: Meta-analysis of Results from Three Studies. PLoS Med 13, e1002146
 (2016).
- 1053 48. Conrad, K.P. Unveiling the vasodilatory actions and mechanisms of relaxin. 1054 *Hypertension* **56**, 2-9 (2010).
- Sun, H.J. *et al.* Relaxin in paraventricular nucleus contributes to sympathetic
 overdrive and hypertension via PI3K-Akt pathway. *Neuropharmacology* **103**, 247-56
 (2016).
- 1058 50. Miyamoto, Y. *et al.* Phosphatidylinositol 3-kinase inhibition induces vasodilator effect of sevoflurane via reduction of Rho kinase activity. *Life Sci* **177**, 20-26 (2017).
- 1060 51. Pawlak, J.B., Wetzel-Strong, S.E., Dunn, M.K. & Caron, K.M. Cardiovascular effects of exogenous adrenomedullin and CGRP in Ramp and Calcrl deficient mice. *Peptides* **88**, 1-7 (2017).
- 1063 52. Ohtsu, H. *et al.* Signal-crosstalk between Rho/ROCK and c-Jun NH2-terminal kinase mediates migration of vascular smooth muscle cells stimulated by angiotensin II.

 1065 *Arterioscler Thromb Vasc Biol* **25**, 1831-6 (2005).
- Tzoulaki, I., Elliott, P., Kontis, V. & Ezzati, M. Worldwide Exposures to Cardiovascular
 Risk Factors and Associated Health Effects: Current Knowledge and Data Gaps.
 Circulation 133, 2314-33 (2016).
- 1069 54. Munafo, M.R., Tilling, K., Taylor, A.E., Evans, D.M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* **47**, 226-235 (2017).
- 1072 55. Pazoki, R. et al. Genetic predisposition to high blood pressure and lifestyle factors:
 1073 Associations with midlife blood pressure levels and cardiovascular events. *Circulation*1074 **137**, 653-661 (2018)
- 1075 56. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An expanded view of complex traits. From polygenic to omnigenic. *Cell* **169**, 1177-1186 (2017)

1078 Figure Legends

- 1079 Figure 1. Study design schematic for discovery and validation of loci. ICBP; International
- 1080 Consortium for Blood Pressure; N, sample size; QC, quality control; PCA, principal-component
- analysis; GWAS, Genome-wide Association Study; 1000G 1000 Genomes; HRC, Haplotype Reference
- Panel; BP: blood pressure; SNPs, single nucleotide polymorphisms; BMI, body mass index; LMM;
- 1083 linear mixed model; UKB, UK Biobank, MAF, minor allele frequency; HLA, Human Leukocyte Antigen;
- 1084 MVP, Million Veterans Program; EGCUT; Estonian Genome Center, University of Tartu; SBP, systolic
- blood pressure; DBP, diastolic blood pressure; PP, pulse pressure.

1086 Figure 2. Manhattan plot showing the minimum P-value for the association across all blood 1087 pressure traits in the discovery stage excluding known and previously reported variants. 1088 Manhattan plot of the discovery genome-wide association meta-analysis in 757,601 individuals 1089 excluding variants in 274 known loci. The minimum P-value, computed using inverse variance fixed 1090 effects meta-analysis, across SBP, DBP and PP is presented. The y axis shows the -log₁₀ P values and 1091 the x axis shows their chromosomal positions. Horizontal red and blue line represents the thresholds 1092 of $P = 5 \times 10^{-8}$ for genome-wide significance and $P = 1 \times 10^{-6}$ for selecting SNPs for replication, respectively. SNPs in blue are in LD ($r^2 > 0.8$) with the 325 novel variants independently replicated 1093 1094 from the 2-stage design whereas SNPs in red are in LD ($r^2 > 0.8$) with 210 SNPs identified through the 1095 1-stage design with internal replication. Any loci in black or grey that exceed the significance 1096 thresholds were significant in the discovery meta-analysis, but did not meet the criteria of

replication in the one- or two-stage designs.

- 1098 Figure 3: Venn Diagrams of Novel Loci Results (a) "Comparison of 1-stage and 2-stage design 1099 analysis criteria": For all 535 novel loci, we compare the results according to the association criteria 1100 used for the one-stage and the two-stage design. Two-hundred and ten loci exclusively met the onestage analysis criteria ($P < 5x10^{-9}$ in the discovery meta-analysis [N=757,601], P < 0.01 in UKB 1101 1102 [N=458,577], P < 0.01 in ICBP [N=299,024] and concordant direction of effect between UKB and 1103 ICBP). The P-values for the discovery and the ICBP meta-analyses were calculated using inverse 1104 variance fixed effects meta-analysis. The P-values in UKB were derived from linear mixed modeling 1105 using BOLT-LMM. Of the 325 novel replicated loci from the 2-stage analysis (genome-wide 1106 significance in the combined meta-analysis, P < 0.01 in the replication meta-analysis and concordant 1107 direction of effect), 204 loci would also have met the one-stage criteria, whereas 121 were only 1108 identified by the two-stage analysis. (b) "Overlap of Associations across Blood Pressure Traits". 1109 For all 535 novel loci, we show the blood pressure traits associated with each locus. We present the 1110 two-stage loci first, followed by the one-stage loci. The locus names provided in alphabetical order 1111 correspond to the nearest annotated gene. SNPs: Single nucleotide polymorphisms; SBP: systolic 1112 blood pressure; DBP: diastolic blood pressure; PP: pulse pressure; UKB: UK Biobank; ICBP: 1113 International Consortium of Blood Pressure.
- 1114 Figure 4. Association of blood pressure loci with lifestyle traits. Plot shows unsupervised 1115 hierarchical clustering of BP loci based on associations with lifestyle-related factors. For the sentinel 1116 SNP at each BP locus (x-axis), we calculated the $-\log_{10}(P)*sign(\beta)$ (aligned to BP-raising allele) as 1117 retrieved from the Gene Atlas catalogue (http://geneatlas.roslin.ed.ac.uk). The P-values in Gene 1118 Atlas were calculated applying linear mixed models. BP loci and traits were clustered according to 1119 the Euclidean distance amongst $-\log_{10}(P)$ *sign(β). Red squares indicate direct associations with the 1120 trait of interest and blue squares inverse associations. Only SNPs with at least one association at P 1121 $<10^{-6}$ with at least one of the traits examined are annotated in the heat-map. All 901 loci are 1122 considered, both known and novel: novel loci are printed in bold font. SNPs: Single Nucleotide 1123 Polymorphisms; BP: Blood Pressure.
- Figure 5. Association of blood pressure loci with other traits. Plot shows results from associations with other traits which were extracted from the GWAS catalog and PhenoScanner databases for the 535 novel sentinel SNPs including proxies in Linkage Disequilibrium (r² ≥ 0.8) with genome-wide significant associations. SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; PP: Pulse Pressure; HR: Heart Rate; ECG: Electrocardiographic traits; CAD: Coronary Artery Disease CHD; Coronary Heart Disease MI; Myocardial Infraction; T2D: Type II Diabetes.
- Figure 6. Association of blood pressure loci with other traits. Plots (a) and (b) show overlap between variants associated to (a) traits and (b) diseases in the manually-curated version of the

DisGeNET database, and all variants in LD r²>0.8 with the known (red bars) SNPs from the 274 1132 1133 published loci, and all (green bars) BP variants from all 901 loci. Numbers on top of the bars denote 1134 the number of SNPs included in DisGeNET for the specific trait or disease. Traits/diseases with an 1135 overlap of at least 5 variants in LD with all markers are shown. The Y axis shows the percentage of 1136 variants associated with the diseases that is covered by the overlap. For the sake of clarity, the 1137 DisGeNET terms for blood pressure and hypertension are not displayed, whereas the following 1138 diseases have been combined: coronary artery disease (CAD), coronary heart disease (CHD) and 1139 myocardial infarction (MI); prostate and breast carcinoma; Crohn's and inflammatory bowel 1140 diseases. 1141 Figure 7. Relationship of deciles of the genetic risk score (GRS) based on all 901 loci with blood 1142 pressure, risk of hypertension and cardiovascular disease in UK Biobank. The plots show sex-1143 adjusted (a) mean systolic blood pressure (SBP) and odds ratios of hypertension (HTN) (N=364,520) 1144 and (b) odds ratios of incident cardiovascular disease (CVD), myocardial infarction (MI) and stroke 1145 (N=392,092), comparing each of the upper nine GRS deciles with the lowest decile; dotted lines 1146 represent the upper 95% confidence intervals. 1147 Figure 8: Known and novel BP associations in the TGFβ signalling pathway. Genes with known 1148 associations with BP are indicated in cyan. Genes with novel associations with BP reported in this 1149 study are indicated in red. TGFB pathway was derived from an ingenuity canonical pathway. BP: 1150 Blood Pressure. 1151

1152 **ONLINE METHODS**

1153 UK Biobank (UKB) data

- 1154 We performed a Genome Wide Association Study (GWAS) analysis in 458,577 UKB
- participants¹³ (**Supplementary Methods**). These consist of 408,951 individuals from UKB
- 1156 genotyped at 825,927 variants with a custom Affymetrix UK Biobank Axiom Array chip and
- 49,626 individuals genotyped at 807,411 variants with a custom Affymetrix UK BiLEVE
- 1158 Axiom Array chip from the UK BiLEVE study⁵⁷, which is a subset of UKB. SNPs were imputed
- centrally by UKB using a reference panel that merged the UK10K and 1000 Genomes Phase
- 1160 3 panel as well as the Haplotype Reference Consortium (HRC) panel⁵⁸. For current analysis
- only SNPs imputed from the HRC panel were considered.
- 1162 UKB phenotypic data
- 1163 Following Quality Control (QC) (Supplementary Methods), we restricted our data to a
- subset of post-QC individuals of European ancestry combining information from self-
- reported and genetic data (Supplementary Methods) resulting in a maximum of N=458,577
- individuals (Fig. 1, Supplementary Fig. 11).
- 1167 Three BP traits were analysed: systolic (SBP), diastolic (DBP) and pulse pressure (PP)
- 1168 (difference between SBP and DBP). We calculated the mean SBP and DBP values from two
- automated (N=418,755) or two manual (N=25,888) BP measurements. For individuals with
- one manual and one automated BP measurement (N=13,521), we used the mean of these
- two values. For individuals with only one available BP measurement (N=413), we used this
- 1172 single value. After calculating BP values, we adjusted for medication use by adding 15 and
- 1173 10 mmHg to SBP and DBP, respectively, for individuals reported to be taking BP-lowering
- medication (N=94,289)⁵⁹. Descriptive summary statistics are shown in **Supplementary Table**
- 1175 **1a**.
- 1176 UKB analysis models
- 1177 For the UKB GWAS we performed linear mixed model (LMM) association testing under an
- additive genetic model of the three (untransformed) continuous, medication-adjusted BP
- traits (SBP, DBP, PP) for all measured and imputed genetic variants in dosage format using
- the BOLT-LMM (v2.3) software¹⁷. We also calculated the estimated SNP-wide heritability
- (h²) in our data. Within the association analysis, we adjust for the following covariates: sex,
- age, age², BMI and a binary indicator variable for UKB vs UK BiLEVE to account for the
- different genotyping chips. The analysis of all HRC-imputed SNPs was restricted to variants
- 1184 with MAF \geq 1% and INFO > 0.1.
- 1185 Genomic inflation and confounding
- We applied the univariate LD score regression method (LDSR)¹⁸ to test for genomic inflation
- 1187 (expected for polygenic traits like BP, with large sample sizes, and especially also from
- analyses of such dense genetic data with many SNPs in high LD)⁶⁰. LDSR intercepts (and

- 1189 standard errors) were 1.217 (0.018), 1.219 (0.020) and 1.185 (0.017) for SBP, DBP and PP
- 1190 respectively, and were used to adjust the UKB GWAS results for genomic inflation, prior to
- the meta-analysis.

1207

International Consortium for Blood Pressure (ICBP) GWAS

- 1193 ICBP GWAS is an international consortium to investigate BP genetics⁶. We combined
- previously reported post-QC GWAS data from 54 studies (N=150,134)^{11,12,61}, with newly
- available GWAS data from a further 23 independent studies (N=148,890) using a fixed
- effects inverse variance weighted meta-analysis. The 23 studies providing new data were:
- 1197 ASCOT-SC, ASCOT-UK, BRIGHT, Dijon 3C, EPIC-CVD, GAPP, HCS, GS:SFHS, Lifelines, JUPITER,
- 1198 PREVEND, TWINSUK, GWAS-Fenland, InterAct-GWAS, OMICS-EPIC, OMICS-Fenland, UKHLS,
- 1199 GoDARTS-Illumina and GoDarts-Affymetrix, NEO, MDC, SardiNIA, METSIM.
- 1200 All study participants were Europeans and were imputed to either the 1000 Genomes
- 1201 Project Phase 1 integrated release v.3 [March 2012] all ancestry reference panel⁶² or the
- HRC panel¹⁶. The final enlarged ICBP GWAS dataset included 77 cohorts (N=299,024).
- 1203 Full study names, cohort information and general study methods are included in
- 1204 **Supplementary Table 1b** and in **Supplementary Tables 20a-c**. GC was applied at study-level.
- 1205 The LDSR intercepts (standard error) for the ICBP GWAS meta-analysis were 1.089 (0.012),
- 1206 1.086 (0.012) and 1.066 (0.011) for SBP, DBP and PP, respectively.

Meta-analyses of discovery datasets

- We performed a fixed-effects inverse variance weighted meta-analysis using METAL^{20,63} to
- 1209 obtain summary results from the UKB and ICBP GWAS, for up to N=757,601 participants and
- 1210 ~7.1 M SNPs with MAF ≥ 1% for variants present in both the UKB data and ICBP meta-
- analysis for all three traits. The LDSR intercepts (standard error), in the discovery meta-
- analysis of UKB and ICBP were 1.156 (0.020), 1.160 (0.021) and 1.113 (0.018) for SBP, DBP
- and PP respectively. The LDSR intercept (standard error), after the exclusion of all published
- 1214 BP variants (see below) in the discovery meta-analysis of UKB and ICBP was 1.090 (0.018),
- 1215 1.097 (0.017) and 1.064 (0.015) for SBP, DBP and PP respectively, hence showing little
- inflation in the discovery GWAS after the exclusion of published loci (Supplementary Fig.
- 1217 **12).** No further correction was applied to the discovery meta-analysis of UKB and ICBP
- 1218 GWAS.

1219

Previously reported variants

- We compiled from the peer-reviewed literature all 357 SNPs previously reported to be
- associated with BP at the time that our analysis was completed, that have been identified
- 1222 and validated as the sentinel SNP in primary analyses from previous BP genetic association
- 1223 studies. These 357 published SNPs correspond to 274 distinct loci, according to locus
- 1224 definition of: (i) SNPs within ±500kb distance of each other; (ii) SNPs in Linkage
- Disequilibrium (LD), using a threshold of $r^2 \ge 0.1$, calculated with PLINK (v2.0). We then

- 1226 augment this list to all SNPs present within our data, which are contained within these 274
- published BP loci, i.e. all SNPs which are located ±500kb from each of the 357 published
- SNPs and/or in LD with any of the 357 previously validated SNPs ($r^2 \ge 0.1$).

1229 Identification of novel signals: Two-stage and one-stage study designs

- 1230 To identify novel signals of association with BP, two complementary study designs (which
- we term here "two-stage design" and "one-stage design") were implemented in order to
- maximize the available data and minimize reporting of false positive associations.

1233 Two-stage design: Overview:

- All of the following criteria had to be satisfied for a signal to be reported as a novel signal of
- association with BP using our two-stage design:
- the sentinel SNP shows significance ($P < 1 \times 10^{-6}$) in the discovery meta-analysis of UKB and ICBP, with concordant direction of effect between UKB and ICBP;
- the sentinel SNP is genome-wide significant ($P < 5 \times 10^{-8}$) in the combined metaanalysis of discovery and replication (MVP and EGCUT) (replication, described below);
- the sentinel SNP shows support (P < 0.01) in the replication meta-analysis of MVP and EGCUT alone (Supplementary Methods);
- 1243 (iv) the sentinel SNP has concordant direction of effect between the discovery and the replication meta-analyses;
- 1245 (v) the sentinel SNP must not be located within any of the 274 previously reported loci described above.
- 1247 The primary replicated trait was then defined as the BP trait with the most significant
- 1248 association from the combined meta-analysis of discovery and replication (in the case
- where a SNP was replicated for more than one BP trait.)

1250 Two-stage design: Selection of variants from the discovery meta-analysis

- 1251 We considered for follow-up SNPs in loci non-overlapping with previously reported loci
- according to both an LD threshold at r² of 0.1 and a 1Mb interval region, as calculated by
- 1253 PLINK⁶⁴. We obtained a list of such SNPs with $P < 1 \times 10^{-6}$ for any of the three BP traits,
- which also had concordant direction of effect between UKB vs ICBP (Supplementary Table
- 1255 **21)**. By ranking the SNPs by significance in order of minimum P-value across all BP traits, we
- performed an iterative algorithm to determine the number of novel signals (Supplementary
- 1257 **Methods),** and identify the sentinel SNP (most significant) per locus.

1258 Two-stage design: Replication analysis

- 1259 We considered SNPs with MAF ≥ 1% for an independent replication in MVP (max
- 1260 N=220,520)¹⁴ and in EGCUT Biobank (N=28,742)¹⁵ (Supplementary Methods). This provides
- 1261 a total of N=249,262 independent samples of European descent available for replication.

- 1262 Additional information on the analyses of the two replication datasets is provided in
- 1263 Supplementary Methods and in Supplementary Table 1c.
- 1264 The two datasets were then combined using fixed effects inverse variance weighted meta-
- analysis and summary results for all traits were obtained for the replication meta-analysis
- 1266 dataset.
- 1267 Two-stage design: Combined meta-analysis of discovery and replication meta-analyses
- 1268 The meta-analyses were performed within METAL software⁶³ using fixed effects inverse
- 1269 variance weighted meta-analysis (Supplementary Methods). The variants from the
- discovery GWAS that required proxies for replication are shown in **Supplementary Table 22**.
- 1271 The combined meta-analysis of both the discovery data (N=757,601) and replication meta-
- analysis (max N=249,262) provided a maximum sample size of N=1,006,863.
- 1273 One-stage design: Overview
- 1274 Variants that were looked-up but did not replicate according to the two-stage criteria were
- 1275 considered in a one-stage design. All of the following criteria had to be satisfied for a signal
- to be reported as a novel signal of association with BP using our one-stage criteria:
- 1277 i) the sentinel SNP has $P < 5 \times 10^{-9}$ in the discovery (UKB+ICBP) meta-analysis;
- the sentinel SNP shows support (P < 0.01) in the UKB GWAS alone;
- the sentinel SNP shows support (P < 0.01) in the ICBP GWAS alone;
- 1280 iv) the sentinel SNP has concordant direction of effect between UKB and ICBP datasets;
- 1282 v) The sentinel SNP must not be located within any of the 274 previously reported loci described above (Supplementary Table 4) or the recently reported non-
- replicated loci from Hoffman et al⁹ (**Supplementary Table 23**).
- 1285 We selected the one-stage *P*-value threshold to be an order of magnitude more stringent
- than a genome-wide significance *P*-value, so as to ensure robust results and to minimize
- 1287 false positive findings. The threshold of $P < 5 \times 10^{-9}$ has been proposed as a more
- 1288 conservative statistical significance threshold, e.g. for whole-genome sequencing-based
- 1289 studies²¹.
- 1290 Selection of variants from the meta-analysis of UKB and ICBP was performed as described
- 1291 above for the two-stage design.
- 1292 **Conditional Analysis**
- 1293 We performed conditional analyses using the GWAS discovery meta-analysis data, in order
- to identify any independent secondary signals in addition to the sentinel SNPs at the 901
- 1295 loci. We used two different methodological approaches, each using the Genome-wide
- 1296 Complex Traits Analysis (GCTA) software²²: (i) full "genome-wide conditional analysis" with
- joint multivariate analysis and stepwise model selection across all three BP traits; and (ii)
- 1298 "locus-specific conditional analysis" for the primary BP trait conditioning on the sentinel

SNPs within each locus (**Supplementary Methods**). For robustness, secondary signals are only reported if obtained from both approaches. All secondary signals were selected at genome-wide significance level, with MAF \geq 1% and confirmed to be pairwise-LD-independent ($r^2 < 0.1$), as well as not being in LD with any of the published or sentinel SNPs at any of the 901 BP-associated loci ($r^2 < 0.1$). In all cases the UKB data was used as the reference genetic data for LD calculation, restricted to individuals of European ancestry only.

Functional analyses: Variants

- 1307 We used an integrative bioinformatics approach to collate functional annotation at both the
- variant level (for each sentinel SNP within all BP loci) and the gene level (using SNPs in LD r²
- 1309 ≥ 0.8 with the sentinel SNPs). At the variant level, we use Variant Effect Predictor (VEP) to
- $1310 \qquad \text{obtain comprehensive characterization of variants, including consequence (e.g. downstream} \\$
- or non-coding transcript exon), information on nearest genomic features and, where
- applicable, amino acid substitution functional impact, based on SIFT and PolyPhen. The
- biomaRt R package is used to further annotate the nearest genes.
- We evaluated all SNPs in LD ($r^2 \ge 0.8$) with our novel sentinel SNPs for evidence of mediation
- of expression quantitative trait loci (eQTL) in all 44 tissues using the Genotype-Tissue
- 1316 Expression (GTEx) database, to highlight specific tissue types which show eQTLs for a larger
- 1317 than expected proportion of novel loci. We further seek to identify novel loci with the
- 1318 strongest evidence of eQTL associations in arterial tissue, in particular. A locus is annotated
- 1319 with a given eGene only if the most significant eQTL SNP for the given eGene is in high LD (r^2
- 1320 ≥ 0.8) with the sentinel SNP, suggesting that the eQTL signal co-localises with the sentinel
- 1321 SNP.

- We annotated nearest genes, eGenes (genes whose expression is affected by eQTLs) and Hi-
- 1323 C interactors with HUVEC, HVMSC and HAEC expression from the Fantom5 project. Genes
- that had higher than median expression levels in the given cell types were indicated as
- 1325 expressed.
- 1326 To identify SNPs in the novel loci that have a non-coding functional effect (influence binding
- 1327 of transcription factors or RNA polymerase, or influence DNase hypersensitivity sites or
- histone modifications), we used DeepSEA, a deep learning algorithm, that learnt the binding
- and modification patterns of ~900 cell/factor combinations⁶⁵. A change of >0.1 in the
- 1330 binding score predicted by DeepSEA for the reference and alternative alleles respectively
- 1331 was used as cut-off to find alleles with non-coding functional effect (Supplementary
- 1332 Methods)
- 1333 We identified potential target genes of regulatory SNPs using long-range chromatin
- interaction (Hi-C) data from HUVECs²³, aorta, adrenal glands, neural progenitor and
- 1335 mesenchymal stem cell, which are tissues and cell types that are considered relevant for
- regulating BP²⁴. We find the most significant promoter interactions for all potential

- regulatory SNPs (RegulomeDB score \leq 5) in LD ($r^2 \geq 0.8$) with our novel sentinel SNPs and
- published SNPs, and choose the interactors with the SNPs of highest regulatory potential to
- 1339 annotate the loci.
- We then performed overall enrichment testing across all loci. Firstly, we used DEPICT⁶⁶
- 1341 (Data-driven Expression Prioritized Integration for Complex Traits) to identify tissues and
- 1342 cells which are highly expressed at genes within the BP loci (Supplementary Methods).
- 1343 Secondly, we used DEPICT to test for enrichment in gene sets associated with biological
- annotations (manually curated and molecular pathways, phenotype data from mouse KO
- 1345 studies) (Supplementary Methods). We report significant enrichments with a false
- discovery rate <0.01. The variants tested were i) the 357 published BP associated SNPs at
- the time of analysis and ii) a set including all (published and novel) variants (with novel SNPs
- 1348 filtered by highest significance, $P < 1 \times 10^{-12}$).
- 1349 Furthermore, to investigate cell type specific enrichment within DNase I sites, we used
- 1350 FORGE, which tests for enrichment of SNPs within DNase I sites in 123 cell types from the
- 1351 Epigenomics Roadmap Project and ENCODE²⁵ (Supplementary Methods). Two analyses
- were compared (i) using published SNPs only; (ii) using sentinel SNPs at all 901 loci, in order
- to evaluate the overall tissue specific enrichment of BP associated variants.

Functional analyses: Genes

1354

1363

1367

- 1355 At the gene level, we used Ingenuity Pathway Analysis (IPA) software (IPA®, QIAGEN
- 1356 Redwood City) to review genes with prior links to BP, based on annotation with the
- "Disorder of Blood Pressure", "Endothelial Development" and "Vascular Disease" Medline
- 1358 Subject Heading (MESH) terms. We used the Mouse Genome Informatics (MGI) tool to
- identify BP and cardiovascular relevant mouse knockout phenotypes for all genes linked to
- 1360 BP in our study. We also used IPA to identify genes that interact with known targets of anti-
- hypertensive drugs. Genes were also evaluated for evidence of small molecule druggability
- or known drugs based on queries of the Drug Gene Interaction database.

Lookups in non-European ancestries

- 1364 As a secondary analysis, we look up all known and novel BP-associated SNPs in Africans
- 1365 (7,782) and South Asians (10,322) from UKB using BOLT-LMM analysis for each BP trait
- within each ancestry (Supplementary Methods).

Effects on other traits and diseases

- We queried SNPs against GWAS catalog²⁶ and PhenoScanner²⁷, including genetics and
- metabolomics databases, to investigate cross-trait effects, extracting all association results
- with genome-wide significance at $P < 5 \times 10^{-8}$ for all SNPs in high LD ($r^2 \ge 0.8$) with the 535
- sentinel novel SNPs, to highlight the loci with strongest evidence of association with other
- 1372 traits. We further evaluated these effects using DisGeNET^{28,29}. At the gene level,
- overrepresentation enrichment analysis (ORA) with WebGestalt⁶⁷ on the nearest genes to
- all BP loci was carried out. Moreover, we tested sentinel SNPs at all published and novel

- 1375 (N=901) loci for association with lifestyle related data including food, water and alcohol
- intake, anthropomorphic traits and urinary sodium, potassium and creatinine excretion
- using the recently developed Stanford Global Biobank Engine and the Gene ATLAS⁶⁸. Both
- are search engines for GWAS findings for multiple phenotypes in UK Biobank. We used a
- Bonferroni corrected significance threshold of $P < 1 \times 10^{-6}$ to deem significance.

Genetic risk scores and percentage of variance explained

- 1381 We calculated a weighted genetic risk score (GRS) (Supplementary Table 24) to provide an
- estimate of the combined effect of the BP raising variants on BP and risk of hypertension
- and applied this to the UKB data (Supplementary Methods). Our analysis included 423,713
- 1384 unrelated individuals of European ancestry of whom 392,092 individuals were free of
- 1385 cardiovascular events at baseline.

1380

- We assessed the association of the continuous GRS variable on BP and with the risk of
- 1387 hypertension, with and without adjustment for sex. We then compared BP levels and risk of
- 1388 hypertension, respectively, for individuals in the top vs bottom guintiles of the GRS
- 1389 distribution. Similar analyses were performed for the top vs bottom deciles of the GRS
- distribution. All analyses were restricted to the 392,092 unrelated individuals of European
- ancestry from UKB. As a sensitivity analysis to assess for evidence of bias in the UKB results,
- 1392 we also carried out similar analyses in Airwave, an independent cohort of N=14,004
- unrelated participants of European descent³⁰ (Supplementary Methods).
- 1394 We calculated the association of the GRS with cardiovascular disease in unrelated
- 1395 participants in UKB data, based on self-reported medical history, and linkage to
- 1396 hospitalization and mortality data (Supplementary Table 25). We use logistic regression
- 1397 with binary outcome variables for composite incident cardiovascular disease
- 1398 (Supplementary Methods), incident myocardial infarction and incident stroke (using the
- 1399 algorithmic UKB definitions) and GRS as explanatory variable (with and without sex
- 1400 adjustment).

1410

- 1401 We also assessed the association of this GRS with BP in unrelated individuals Africans
- 1402 (N=6,970) and South Asians (N=8,827) from the UKB to see whether BP-associated SNPs
- 1403 identified from GWAS predominantly in Europeans are also associated with BP in
- 1404 populations of non-European ancestry.
- 1405 We calculated the percentage of variance in BP explained by genetic variants using the
- 1406 independent Airwave cohort (N=14,004) (Supplementary Methods). We considered three
- different levels of the GRS: (i) all pairwise-independent, LD-filtered ($r^2 < 0.1$) published SNPs
- 1408 within the known loci; (ii) all known SNPs and sentinel SNPs at novel loci; (iii) all
- 1409 independent signals at all 901 known and novel loci including the 163 secondary SNPs.

Data availability statement

- 1411 The UKB GWAS data can be assessed from the UK Biobank data repository
- 1412 (http://biota.osc.ox.ac.uk/). The genetic and phenotypic UKB data are available upon

- application to the UK Biobank (https://www.ukbiobank.ac.uk). ICBP summary data can be
- 1414 assessed through request to ICBP steering committee. Contact Mark Caulfield
- 1415 (m.j.caulfield@qmul.ac.uk) or Paul Elliott (p.elliott@imperial.ac.uk) to apply for access to
- 1416 the data. The UKB+ICBP summary data can be assessed through request to Paul Elliott
- 1417 (p.elliott@imperial.ac.uk) or Mark Caulfield (m.j.caulfield@qmul.ac.uk). All replication data
- 1418 generated during this study are included in the published article. For example, association
- results of look-up variants from our replication analyses and the subsequent combined
- meta-analyses are contained within the Supplementary Tables provided.

1421 Reporting Summary

- 1422 Further information on experimental design is available in the Life Sciences Reporting
- 1423 Summary linked to this article.

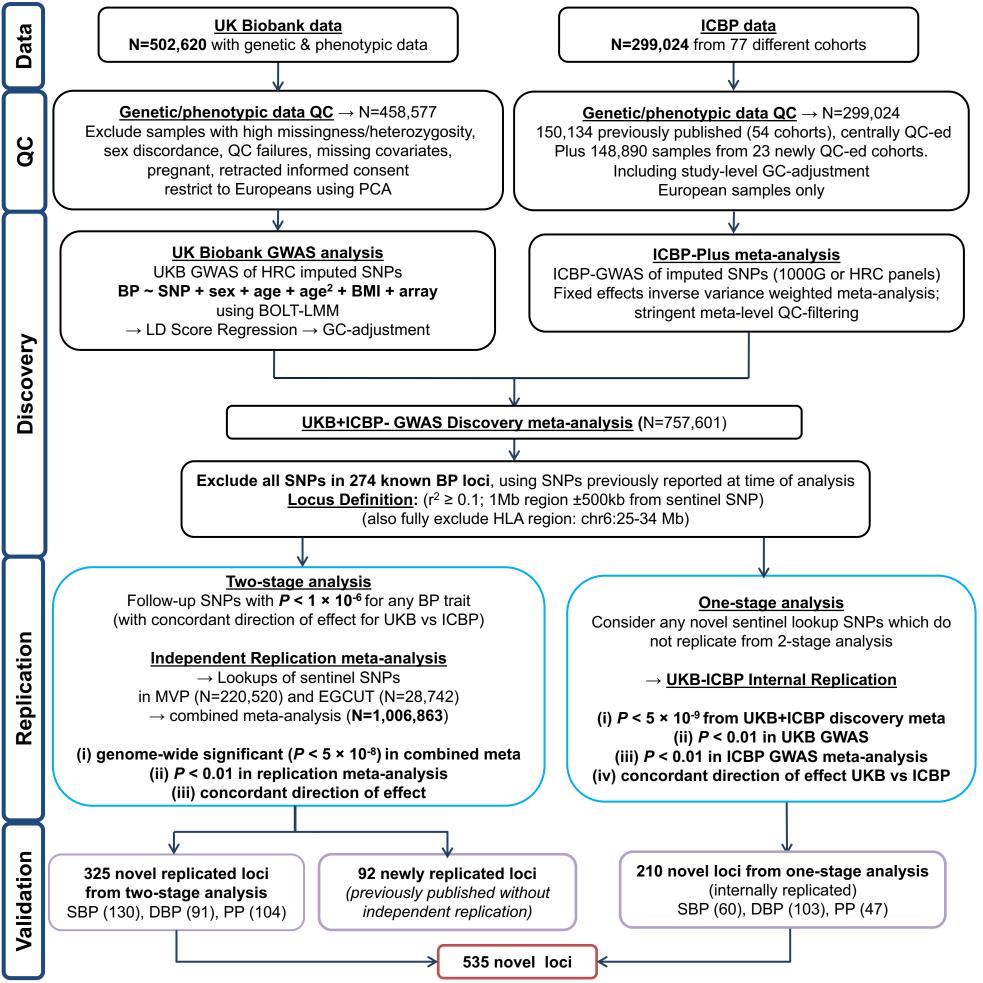
1424 Ethics Statement

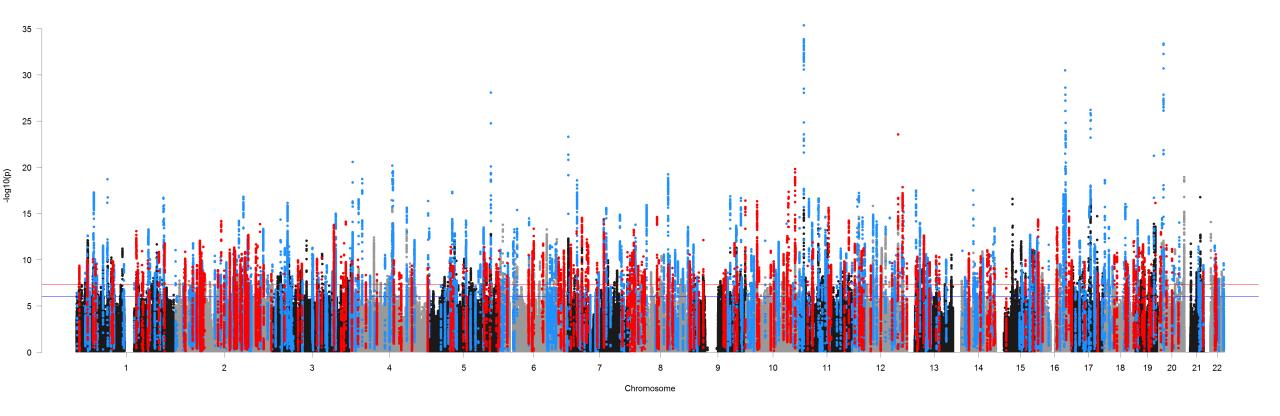
- 1425 The UKB study has approval from the North West Multi-Centre Research Ethics Committee.
- 1426 Any participants from UKB who withdrew consent have been removed from our analysis.
- 1427 Each cohort within the ICBP meta-analysis as well as our independent replication cohorts of
- 1428 MVP and EGCUT had ethical approval locally. More information on the participating cohorts
- is available in **Supplementary Methods**.

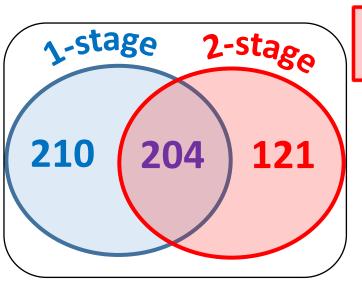
1430 References

- 1431 57. Wain, L.V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study
- in UK Biobank. *Lancet Respir Med* **3**, 769-81 (2015).
- 1434 58. Bycroft, C.F., C; Petkova, D; Band, G; Elliot, LT; Sharp, K; Motyer, A; Vukcevic, D;
 - Delaneau, O; O'Conell, J; Cortes, A; Welsh, S; McVean, G; Leslie, S; Donelly, P;
- 1436 Marchini, J. Genome-wide geentic data on 500,000 UK Biobank Participants. *bioRxiv* 1437 **166298** (2017).
- 1438 59. Tobin, M.D., Sheehan, N.A., Scurrah, K.J. & Burton, P.R. Adjusting for treatment
- effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* **24**, 2911-35 (2005).
- 1441 60. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. 1442 *Nature* **542**, 186-190 (2017).
- Wain, L.V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* **43**, 1005-11 (2011).
- 1445 62. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
- Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
- 1449 64. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-1450 based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
- 1451 65. Zhou, J. & Troyanskaya, O.G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* **12**, 931-4 (2015).
- 1453 66. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).

1455	67.	Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more
1456		comprehensive, powerful, flexible and interactive gene set enrichment analysis
1457		toolkit. Nucleic Acids Res (2017).
1458	68.	Canela-Xandri, O.R., Konrad; Tenesa, Albert. An atlas of genetic associations in UK
1459		Biobank. bioRxiv 176834 (2017).
1460		







(65 2-stage): ADORA1, APOLD1, ARHGAP29, ARIH2, ARL14EP, BCAR3, BCAS3, C11orf24, C1orf172, CCT6A, CITED2, CNTN3, CTC-228N24.3, DENND2A, DGKH, ERBB4, FAM193A, FAM208B, FOXC1, GABRA2, GDF2, GLIS3, HSPA12A, IER5L, IRF6, JAZF1, KAT2B, LINCO0311, MCM9, MERTK, MLF1, NRXN1, PCCB, PLXNB2, POM121C, PRKD1, RARRES2, RBFOX1, RBM26, RBMS1, RNASEH2B, RP11-1055B8.6, RTN4, SEMA4A, SKI, SLC30A5, SOX5, SPIB, SREK1, ST5, SYT1, SZT2, TARS, TFCP2L1, THSD7B, TMEM108, TNKS, TOX, TRIP12, WDR7, WNT4, XPR1, ZBTB20, ZNF804A, ZSWIM2

(32 1-stage): AC009120.4, AC010967.2, AGBL4, AHRR, APOH, BANP, CAND1, CBWD1, CLDN23, CLN8, DFNA5, DMRTA1, FBRSL1, FGR, FOXF1, FOXO6, GRIN2B, HTRA1, KANK1, LINCO1091, LRBA, OLA1, PDE11A, PEPD, PKN2, PREX2, PRR20A, RGMB, RP11-122C21.1, RP11-428C19.4, snoU13, WASF3

(57 2-stage): ABCC9, AC004156.3, AC017083.3, AL163953.3, AP000721.4, APOE, ARHGEF25, CDKAL1, CELF2, CENPP, COG5, COL15A1, CRB1, CTNNB1, EBF1, EDN1, EPB41L2, ERAL1, FGF9, FZD2 GIPR, H1FNT, IGF1, INPP5A, KCNQ5 KDM4B, LCORL, LIN54, LRCH1, LTBP2, MEIS1, MSRA, MYO1E, NCOA7, NEK6, NNT, NT5C1B, OSBPL7, PBX3, PDE8A, PHACTR4, PHTF2, PPP2R2D, PRR16, RAD52, RP11-158M2.4, RP11-89M16.1, SLC30A10, TGFBR2, THADA, TRHR, TRIOBP-NOL12, TSPAN14, WHSC1L1, ZCCHC2, ZMAT2, ZNF618 (17 1-stage): CDYL2, CWC27, FAM46A, FBXO33, FOXO3, HHEX, MC4R, RNF130, RP11-227G15.6, SIRT1, SLC4A10, SPATS2L, TNS3, TTC28, UBAP1, WT1, YEATS2

(73 2-stage): AC005027.3, AC011294.3, AC069368.3, AC074391.1, ADAMTSL3, AKR1A1, ALDH8A1, ANKUB1, AP000320.7-AP000318.2, BMP2, BNC2, BRD1, BUD13, C1QTNF7, CCDC30, CDKN1A, CLEC16A, CMSS1, CNOT1, CYBRD1, DDAH1, DIP2A, DIRC1, DYNLRB1, FBXL17, FGD6, FOXD1, G6PC2, HAUS6, HSF2, IL6, ITGA1, ITGA9, KIF15, LCA5L, LIG3, LINCO0521, LINCO0536, LRRC69, MALRD1, MAP2K2, MED13L, MN1, MSI2, MTNR1B, MXRA7, NDUFAF6, ODF2L, PAPPA, PDE3A, PHC2, PSMG2, QSOX1, RAMP2, RN7SKP15, RNF219, RP11-339B21.8, RP11-432I5.2, RREB1, SAMD4A, SCN10A, SGIP1, SHOX2, SMOC2, SNX19, STAM2, TRANK1, UBE2I, UQCR10, WDR1, YAP1, YY1, ZNF385B (41 1-stage): AC007381.2, ADCY5, APOB, BCAT1, C10orf76, CDK14, CHD2, CHRM2, COL6A1, CTC-340123.2, DAZAP1, DIO3, EEPD1, ELL, FHL2, FNDC3B, FOXN3, KLHL29, LIMK1, MAEA, MAK16-TTI2, PPP4R2, PROM1, PRPF40A, RIN3, RN7SL89P, RNF144B, RP11-15B24.5, RP1-130H16.18, RP11-373N22.3, RP11-497E19.2,

DBP

(64 2-stage): AC011518.1, AC073218.1, AKR1B10, ANO1, ATAD5, AUTS2, BEND7, BTBD3, CCKBR, CD160, CDK17, CITED2, COLEC11, CTAGE1, CTBP2, CYP27A1, DNAJB4, EPN2, FOXK1, GRM7, GTF2I, GYPC, HSPA4, IGFBP7, LINCO0211, MBNL2, MIR4421, MMP14, NACA, NCOR2, OR51E1, PCDH17, PDLIM5, PGR, PIAS1, PIEZO2, PIK3R3, PKD2L1, PLEC, POLD3, POLN, PPM1A, PRSS50, RERG, RP11-1038A11.3, RP11-20D14.4, RP11-34N19.1, RP4-655J12.4, RP4-712E4.1, RPS27P25, SCN2A, SHFM1, SLCO3A1, STARD6, STK38L, TMEM44, TRIM13, TRMT10C, UBE2E2, VEGFA, WDR90, ZAP70, ZNF462, ZSCAN2

(53 2-stage): AC022431.2, ACTBL2, ACVR1C, AL672294.1, (8 2-stage): BCL2, KLF14, L2HGDH, AQP1, ARHGAP15, ATP2B1, BANK1, BMPR1B, C12orf75, C1orf21, CAPRIN1, CTD-2260A17.2, CTD-2349B8.1, DLG1, (41-stage): ARMC4, FGFR2, FAM129B, FARP2, IRX1, KIF2A, KLF5, LINCO1006, MAST4, SNRNP70, YTHDF3 MECR, MEF2A, MEX3C, MORC3, NFATC2, NRG4, NUDT3, OPRM1, PDP2, PLA2G12B, PPM1E, PTK2, PTPRD, RGS6, RP11-444A22.1, RP11-455F5.3, RP11-714L20.1, RP11-805L22.1, SNX6,

LRP4, MARK3, PDGFC, RXFP2, TERT SORBS3, STIM2, SWAP70, TGFB2, TMEM107, TOP3A, TRIM48, TSNARE1, Y RNA, YES1, ZEB2, ZFAND2A (28 1-stage): AC005592.2, AC019181.3, AC021218.2, AEBP2, ANK3, AQP4-AS1, ARHGEF26, C16orf97, CLPB, CPS1, EGFL7, ICOS, LA16c-306E5.3, NAA16, NCALD, NKD1, PAX8, RAPGEF5, RBPMS, RNU6-192P, RP11-453O22.1, RP1-74B13.2, SMOX, SORCS3, STXBP5,

(5 2-stage): HAS3, KIAA1755, SLX4IP, STEAP2, UBE2D3

RP11-95P2.1, SLC22A3, SNORA40, TBL1XR1, TBX18, TET1, TGFBR3, TMEM239, ZNF467, ZNF516

TSHZ1, ZEB2, ZFP36L1 (88 1-stage): AC053503.11, AC068196.1, AC083949.1, AC097495.2, ACVR2A, AGPAT4, AP1B1P1, ARAP2, ASXL3, ATP10A, ATP12A, ATXN7, BCKDHB, BTRC, C1GALT1, C9orf170, CACNA1C, CAMTA1, CCDC33, CCDC68, CCM2, CCNT2-AS1, CDK5RAP1, CLNS1A, CTC-360G5.8, DACH1, DCDC1, DGKB, DPYSL2, DUSP1, EPC1, EXOC6B, FAM168A, FLI00388, GAB2, GLI2, GRB10, HDAC4, KANK3, KB-1507C5.2, KCNB1, KIF26A, KLF2, KLF7, LAMC1, LPHN3, MCPH1, MIR3927, MLTK, MLXIP, MRPS31, MXD3, MXI1, NEDD4L, NTN4, OGFRL1, OSGIN2, PFKFB2, PLCXD2, PLEKHH1, PLK2, POU2F1, PPHLN1, PRDM1, PTPRD, RBMS3, RGS17, RP11-125B21.2, RP11-145M9.2, RP11-264C15.2, RP11-302M6.4, RP11-506O24.1, RP11-61G23.1, RP11-65J21.4, SIK2, SLC7A2, SPRY2, SYNPO2, TBX20, TCF4, TDRD5, UBAP2L, UBASH3B, USP34, VPS54, WAC, ZNF100, ZNF423

