Genetic analysis of over one million people identifies 535 novel loci for blood pressure.

Short title: Blood pressure GWAS in one million people

Evangelos Evangelou[1,2]*, Helen R Warren[3,4]*, David Mosen-Ansorena[1]*, Borbala Mifsud[3]*, Raha Pazoki[1]*, He Gao[1,5]*, Georgios Ntritsos[2]*, Niki Dimou[2]*, Claudia P Cabrera[3,4], Ibrahim Karaman[1], Fu Liang Ng[3], Marina Evangelou[1,6], Katarzyna Witkowska[3], Evan Tzanis[3], Jacklyn N Hellwege[7] , Ayush Giri[8], Digna R Velez-Edwards[8], Yan V Sun[9,10],Kelly Cho[11,12], J.Michael Gaziano[11,12], Peter WF Wilson[13], Philip S Tsao[14], Csaba P Kovesdy[15], Tonu Esko[16,17], Reedik Mägi[16], Lili Milani[16], Peter Almgren[18], Thibaud Boutin[19], Stéphanie Debette[20,21], Jun Ding[22], Franco Giulianini[23], Elizabeth G Holliday[24], Anne U Jackson[25], Ruifang Li-Gao[26], Wei-Yu Lin[27], Jian'an Luan[28], Massimo Mangino[29,30], Christopher Oldmeadow[24], Bram Prins[31], Yong Qian[22], Muralidharan Sargurupremraj[21], Nabi Shah[32,33], Praveen Surendran[27], Sébastien Thériault[34,35], Niek Verweij[17,36,37], Sara M Willems[28], Jing-Hua Zhao[28], Philippe Amouyel[38], John Connell[39], Renée de Mutsert[26], Alex SF Doney[32], Martin Farrall[40,41], Cristina Menni[29], Andrew D Morris[42], Raymond Noordam[43], Guillaume Paré[34], Neil R Poulter[44], Denis C Shields[45], Alice Stanton[46], Simon Thom[44], Gonçalo Abecasis[47], Najaf Amin[48], Dan E Arking[49], Kristin L Ayers[50], Caterina M Barbieri[51,52], Chiara Batini[53], Joshua C Bis[54], Tineka Blake[53], Murielle Bochud[55], Michael Boehnke[25], Eric Boerwinkle[56], Dorret I Boomsma[57], Erwin P Bottinger[58], Peter S Braund[59,60], Marco Brumat[61], Archie Campbell[62,63], Harry Campbell[64], Aravinda Chakravarti[49], John C Chambers[1,5,65-67], Ganesh Chauhan[68], Marina Ciullo[69,70], Massimiliano Cocca[52], Francis Collins[71], Heather J Cordell[72], Gail Davies[73,74], Martin H de Borst[75], Eco J de Geus[57], Ian J Deary[73,74], Joris Deelen[76], Fabiola Del Greco M[77], Cumhur Yusuf Demirkale[78], Marcus Dörr[79,80], Georg B Ehret[49,81], Roberto Elosua[82], Stefan Enroth[83], A Mesut Erzurumluoglu[53], Teresa Ferreira[84], Mattias Frånberg[85-87], Oscar H Franco[88], Ilaria Gandin[61], Paolo Gasparini[61,89], Vilmantas Giedraitis[90], Christian Gieger[91-93], Giorgia Girotto[61,94], Anuj Goel[40,41], Alan J Gow[73,95], Vilmundur Gudnason[96,97], Xiuqing Guo[98], Ulf Gyllensten[83], Anders Hamsten[85,86], Tamara B Harris[99], Sarah E Harris[62,73], Catharina A Hartman[100], Aki S Havulinna[101,102], Andrew A Hicks[77], Edith Hofer[103,104], Albert Hofman[88,105], Jouke-Jan Hottenga[57], Jennifer E Huffman[19], Shih-Jen Hwang[106,107], Erik Ingelsson[108,109], Alan James[110,111], Rick Jansen[112], Marjo-Riitta Jarvelin[5,113-115], Roby Joehanes[116,117], Åsa Johansson[83], Andrew D Johnson[117,118], Peter K Joshi[64], Pekka Jousilahti[101], J Wouter Jukema[119], Antti Jula[101], Mika Kähönen[120,121], Sekar Kathiresan[17,36,122], Bernard D Keavney[123,124], Kay-Tee Khaw[125], Paul Knekt[101], Joanne Knight[126], Ivana Kolcic[127], Jaspal S Kooner[5,66,67,128], Seppo Koskinen[101], Kati Kristiansson[101], Zoltan Kutalik[55,129], Maris Laan[130], Marty Larson[117], Lenore J Launer[99], Benjamin C Lehne[1], Terho Lehtimäki[131,132], David CM Liewald[73,74], Li Lin[81], Lars Lind[133], Cecilia M Lindgren[40,84,134], YongMei Liu[135], Ruth JF Loos[28,58,136], Lorna M Lopez[73,137,138], Yingchang Lu[58], Leo-Pekka Lyytikäinen[131,132], Anubha Mahajan[40], Chrysovalanto Mamasoula[139], Jaume Marrugat[82], Jonathan Marten[19], Yuri Milaneschi[140], Anna Morgan[61], Andrew P Morris[40,141], Alanna C Morrison[142], Peter J Munson[78], Mike A Nalls[143,144], Priyanka Nandakumar[49], Christopher P Nelson[59,60], Teemu Niiranen[101,145], Ilja M Nolte[146], Teresa Nutile[69], Albertine J Oldehinkel[147], Ben A Oostra[48], Paul F O'Reilly[148], Elin Org[16], Sandosh Padmanabhan[63,149], Walter Palmas[150], Aarno Palotie[102,151,152], Alison Pattie[74], Brenda WJH Penninx[140], Markus Perola[101,102,153], Annette Peters[92], Ozren Polasek[127,154], Peter P Pramstaller[77,155,156], Nguyen Quang Tri[78], Olli T Raitakari[157,158], Meixia Ren[159,160], Rainer Rettig[161], Kenneth Rice[162], Paul M Ridker[23,163], Janina S Ried[92], Harriëtte Riese[147], Samuli Ripatti[102,164], Antonietta Robino[52], Lynda M Rose[23], Jerome I Rotter[98], Igor Rudan[165], Daniela Ruggiero[69], Yasaman Saba[166], Cinzia F Sala[51], Veikko Salomaa[101], Nilesh J Samani[59,60], Antti-Pekka Sarin[102], Reinhold Schmidt[103], Helena Schmidt[166], Nick Shrine[53], David Siscovick[167], Albert V Smith[96,97], Harold Snieder[146], Siim Sõber[130], Rossella Sorice[69], John M Starr[73,168], David J Stott[169], David P Strachan[170], Rona J Strawbridge[85,86], Johan Sundström[133], Morris A Swertz[171], Kent D Taylor[98], Alexander Teumer[80,172], Martin D Tobin[53], Maciej

Tomaszewski[123,124], Daniela Toniolo[51], Michela Traglia[51], Stella Trompet[119,173], Jaakko Tuomilehto[174-177], Christophe Tzourio[21], André G Uitterlinden[88,178], Ahmad Vaez[146,179], Peter J van der Most[146], Cornelia M van Duijn[48], Anne-Claire Vergnaud[1], Germaine C Verwoert[88], Veronique Vitart[19], Uwe Völker[80,180], Peter Vollenweider[181], Dragana Vuckovic[61,182], Hugh Watkins[40,41], Sarah H Wild[183], Gonneke Willemsen[57], James F Wilson[19,64], Alan F Wright[19], Jie Yao[98] , Tatijana Zemunik[184], Weihua Zhang[1,66], John R Attia[24], Adam S Butterworth[27,185], Daniel I Chasman[23,163], David Conen[186,187], Francesco Cucca[188,189], John Danesh[27,185], Caroline Hayward[19], Joanna MM Howson[27], Markku Laakso[190], Edward G Lakatta[191], Claudia Langenberg[28], Olle Melander[18], Dennis O Mook-Kanamori[26,192], Colin NA Palmer[32], Lorenz Risch[193-195], Robert A Scott[28], Rodney J Scott[24], Peter Sever[128], Tim D Spector[29], Pim van der Harst[196], Nicholas J Wareham[28], Eleftheria Zeggini[31], Daniel Levy[117,118], Patricia B Munroe[3], Christopher Newton-Cheh[134,197,198], Morris J Brown[3,4], Andres Metspalu[16], Adriana M Hung[199], Christopher J O'Donnell[200], Todd L Edwards[7] 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[53]*, Paul Elliott[1,5,203,204]*‡, Mark J Caulfield[3,4]*‡

* Equal contribution

‡ Corresponding authors

1. Department of Epidemiology and Biostatistics, Imperial College London, London, UK.
2. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece.
3. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
4. National Institute for Health Research, Barts Cardiovascular Biomedical Research Center, Queen Mary University of London, London, UK.
5. MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK.
6. Department of Mathematics, Imperial College London, UK
7. Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN.
8. Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center; Tennessee Valley Health Systems VA, Nashville, TN.
9. Department of Epidemiology and Biomedical Informatics, Emory University Rollins School of Public Health, Atlanta GA, USA.
10. Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA.
11. Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130, USA.
12. Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, MA, Department of Medicine, Harvard Medical School, Boston, MA.

13. Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA, 30233.
14. VA Palo Alto Health Care System; Division of Cardiovascular Medicine, Stanford University School of Medicine.
15. Nephrology Section, Memphis VA Medical Center, Memphis TN.
16. Estonian Genome Center, University of Tartu, Tartu, Estonia.
17. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA.
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, EH4 2XU Scotland, UK
20. Department of Neurology, Bordeaux University Hospital, F-33000 Bordeaux, France.
21. Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, CHU Bordeaux, F-33000 Bordeaux, France.
22. Laboratory of Genetics and Genomics,NIA/NIH , Baltimore, MD 21224,USA.
23. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA.
24. Hunter Medical Reseach Institute and Faculty of Health, University of Newcastle, New Lambton Heights, New South Wales, Australia.
25. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA.
26. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.
27. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
28. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0QQ, UK.
29. Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, UK.
30. NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK.
31. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK.
32. Division of Molecular and Clinical Medicine, School of Medicine, University of Dundee, DD1 9SY, UK.
33. Department of Pharmacy, COMSATS Institute of Information Technology, Abbottabad, 22060, Pakistan.
34. Department of Pathology and Molecular Medicine, McMaster University, 1280 Main St W, Hamilton, L8S 4L8, Canada.
35. Institut universitaire de cardiologie et de pneumologie de Québec-Université Laval, 2725 Chemin Ste-Foy, Quebec City, G1V 4G5, Canada.
36. Cardiovascular Research Center and Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA.
37. University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, The Netherlands.
38. University of Lille, Inserm, Centre Hosp. Univ Lille, Institut Pasteur de Lille, UMR1167 - RID-AGE - Risk factors and molecular determinants of aging-related diseases, Epidemiology and Public Health Department, F-59000 Lille, France.

39. University of Dundee, Ninewells Hospital & Medical School, Dundee, DD1 9SY, UK.
40. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
41. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 9DU, UK.
42. Usher Institute of Population Health Sciences and Informatics, University of Edinburgh.
43. Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands.
44. International Centre for Circulatory Health, Imperial College London, London, W2 1PG, UK.
45. School of Medicine, University College Dublin, Ireland.
46. Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland.
47. Center for Statistical Genetics, Dept. of Biostatistics, SPH II, 1420 Washington Heights, Ann Arbor, MI 48109-2029.
48. Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC, Rotterdam, 3000CA, the Netherlands.
49. Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
50. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
51. Division of Genetics and Cell Biology, San Raffaele Scientific Institute, 20132 Milano, Italy.
52. Institute for Maternal and Child Health IRCCS Burlo Garofolo, Via dell'Istria 65, Trieste, 34200, Italy.
53. Department of Health Sciences, University of Leicester, Leicester, UK.
54. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA.
55. Institute of Social and Preventive Medicine, University Hospital of Lausanne, Route de la Corniche 10, 1010 Lausanne, Switzerland.
56. 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, 77030, USA.
57. Department of Biological Psychology, Vrije Universiteit Amsterdam, EMGO+ institute, VU University medical center, Amsterdam, the Netherlands.
58. The Charles Bronfman Institute for Personalized Medicine, Icachn School of Medicine at Mount Sinai, New York, NY 10029.
59. Department of Cardiovascular Sciences, University of Leicester, Leicester LE3 9QP, UK.
60. NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, UK.
61. Department of Medical, Surgical and Health Sciences, University of Trieste, Strada di Fiume 447, Trieste, 34100, Italy.
62. Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.

63. Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, UK.
64. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh EH89AG, Scotland, UK
65. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore.
66. Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK.
67. Imperial College Healthcare NHS Trust, London W12 0HS, UK.
68. Centre for Brain Research, Indian Institute of Science, Bangalore, India.
69. Institute of Genetics and Biophysics "A. Buzzati-Traverso", CNR, via P. Castellino 111, 80131 Napoli - Italy.
70. IRCCS Neuromed, Pozzilli, Isernia, Italy.
71. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA.
72. Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK.
73. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.
74. Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK.
75. Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands.
76. Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, 2300RC, the Netherlands.
77. Institute for Biomedicine, Eurac Research, Bolzano, Italy - Affiliated Institute of the University of Lübeck, Germany.
78. Mathematical and Statistical Computing Laboratory, Office of Intramural Research, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892.
79. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, 17475, Germany.
80. DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, 17475, Germany.
81. Cardiology, Department of Medicine, Geneva University Hospital, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva 14, Switzerland.
82. CIBERCV & Cardiovascular Epidemiology and Genetics, IMIM. Dr Aiguader 88, Barcelona, 08003 Spain.
83. Department of Immunology, Genetics and Pathology, Uppsala Universitet, Science for Life Laboratory, Husargatan 3, Uppsala, SE-75108,Sweden.
84. Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK.
85. Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, 17176, Sweden.
86. Centre for Molecular Medicine, L8:03, Karolinska Universitetsjukhuset, Solna, 171 76, Sweden.
87. Department of Numerical Analysis and Computer Science, Stockholm University, Lindstedtsvägen 3, Stockholm, 100 44, Sweden.
88. Department of Epidemiology, Erasmus MC, Rotterdam, 3000CA, the Netherlands.
89. IRCCS-Burlo Garofalo Children Hospital, Via dell'Istria 65, Trieste, Italy.

90. Department of Public Health and Caring Sciences, Geriatrics, Uppsala 752 37, Sweden.
91. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
92. Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
93. German Center for Diabetes Research (DZD), Neuherberg, Germany.
94. IRCCS-Burlo Garofalo Children Hospital, Via dell'Istria 65, Trieste, Italy.
95. Department of Psychology, School of Social Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK.
96. Faculty of Medicine, University of Iceland, Reykjavik, Iceland.
97. Icelandic Heart Assoication, Kopavogur, Iceland.
98. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, 1124 W. Carson Street, Torrance, CA 90502, USA.
99. Intramural Research Program, Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging.
100. Department of Psychiatry, University of Groningen, University Medical Center Groningen, Hanzeplein 1, PO Box 30001, 9700 RB Groningen, The Netherlands.
101. Department of Public Health Solutions, National Institute for Health and Welfare (THL), Helsinki, Finland.
102. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.
103. Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Auenbruggerplatz 22, 8036 Graz, Austria.
104. Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Auenbruggerplatz 2, 8036 Graz, Austria.
105. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA.
106. The Framingham Heart Study, Framingham MA 01702.
107. The Population Science Branch, Division of Intramural Research, National Heart Lung and Blood Institute national Institute of Health, Bethesda, MD 20892.
108. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala 752 37, Sweden.
109. Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
110. Department of Pulmonary Physiology and Sleep, Sir Charles Gairdner Hospital, Hospital Avenue, Nedlands 6009,H57 Western Australia.
111. School of Medicine and Pharmacology, University of Western Australia.
112. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, the Netherlands.
113. Biocenter Oulu, P.O.Box 5000, Aapistie 5A, FI-90014 University of Oulu, Finland.
114. Center For Life-course Health Research, P.O.Box 5000, FI-90014 University of Oulu, Finland.
115. Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O.Box 20, FI-90220 Oulu, 90029 OYS, Finland.
116. Hebrew SeniorLife, Harvard Medical School, 1200 Centre Street Room #609, Boston, MA 02131.

117. National Heart, Lung and Blood Institute's Framingham Heart Study, Framingham, MA 01702, USA.
118. Population Sciences Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20814, USA.
119. Department of Cardiology, Leiden University Medical Center, Leiden, 2300RC the Netherlands.
120. Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland.
121. Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33014, Finland.
122. Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA.
123. Division of Cardiovascular Sciences, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK.
124. Division of Medicine, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
125. Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge CB2 2SR, UK.
126. Data Science Institute and Lancaster Medical School, LA1 4YG, Lancaster.
127. Department of Public Health, Faculty of Medicine, University of Split, Croatia.
128. National Heart and Lung Institute, Imperial College London, London, UK.
129. Swiss Institute of Bioinformatics, Lausanne, Switzerland.
130. Institute of Biomedicine and Translational Medicine, University of Tartu, Ravila Str. 19, 50412 Tartu, Estonia.
131. Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland.
132. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33014, Finland
133. Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala 751 85, Sweden.
134. Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.
135. Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, 27106, USA.
136. Mindich Child health Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
137. 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, UCD, Belfield, Dublin, Ireland.
139. Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK.
140. Department of Psychiatry, Amsterdam Public Health and Amsterdam Neuroscience, VU University Medical Center/GGZ inGeest, Oldenaller 1, 1081 HJ Amsterdam Amsterdam, The Netherlands.
141. Department of Biostatistics, University of Liverpool, Block F, Waterhouse Building, 1-5 Brownlow Street, Liverpool L69 3GL, UK.

142. Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, 1200 Pressler St., Suite 453E, Houston, TX 77030, USA.
143. Data Tecnica International, Glen Echo, MD, USA.
144. Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA.
145. Department of Medicine, Turku University Hospital and University of Turku, Finland.
146. Department of Epidemiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, PO Box 30001, 9700 RB Groningen, The Netherlands.
147. Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), University of Groningen, University Medical Center Groningen, PO Box 30001 (CC72), 9700 RB Groningen, The Netherlands.
148. SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK.
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 G12 8TA, UK.
150. Department of Medicine, Columbia University Medical Center, 622 West 168th Street, PH 9, East, 107, New York, NY 10032, USA.
151. Analytic and Translational Genetics Unit, Department of Medicine, Department of Neurology and Department of Psychiatry Massachusetts General Hospital, Boston, MA, USA.
152. The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA.
153. University of Tartu, Tartu, Estonia.
154. Psychiatric hospital “Sveti Ivan”, Zagreb, Croatia.
155. Department of Neurology, General Central Hospital, Bolzano, Italy.
156. Department of Neurology, University of Lübeck, Lübeck, Germany.
157. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20521, Finland.
158. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland.
159. Department of Geriatric Medicine, Fujian Provincial Hospital, Fujian Key Laboratory of Geriatrics, Fujian Medical University, Fuzhou, 350001, China.
160. William Harvey Research Institute Barts and The London School of Medicine and Dentistry, Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London EC1M 6BQ, United Kingdom.
161. Institute of Physiology, University Medicine Greifswald, Karlsburg, 17495, Germany.
162. Department of Biostatistics University of Washington, Seattle, WA 98101, USA.
163. Harvard Medical School, Boston MA.
164. Public health, Faculty of Medicine, University of Helsinki, Finland
165. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh EH89AG, Scotland, UK.
166. Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Neue Stiftingtalstraße 6, 8010 Graz, Austria.
167. The New York Academy of Medicine. 1216 5th Ave, New York, NY 10029, USA.

168. Alzheimer Scotland Dementia Research Centre, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK.
169. Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, United Kingdom.
170. Population Health Research Institute, St George's, University of London, London SW17 0RE, UK.
171. Department of Genetics, University of Groningen, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands.
172. Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany.
173. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, 2300RC, the Netherlands.
174. Dasman Diabetes Institute, Dasman, 15462 Kuwait.
175. Chronic Disease Prevention Unit, National Institute for Health and Welfare, 00271 Helsinki, Finland.
176. Department of Public Health, University of Helsinki, 00014 Helsinki, Finland.
177. Saudi Diabetes Research Group, King Abdulaziz University, 21589 Jeddah, Saudi Arabia.
178. Department of Internal Medicine, Erasmus MC, Rotterdam, 3000CA, the Netherlands.
179. Research Institute for Primordial Prevention of Non-communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran.
180. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, 17475, Germany.
181. Department of Internal Medicine, University Hospital, CHUV, 1011 Lausanne, Switzerland.
182. Experimental Genetics Division, Sidra Medical and Research Center, PO Box 26999, Doha, Qatar.
183. Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh EH89AG, Scotland, UK
184. Department of Biology, Faculty of Medicine, University of Split, Croatia.
185. The National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, UK.
186. Division of Cardiology, University Hospital, Petersgraben 4, 4031 Basel, Switzerland.
187. Division of Cardiology, Department of Medicine, McMaster University, 1280 Main St W, Hamilton, L8S 4L8, Canada.
188. Institute of Genetic and Biomedical Research, National Research Council (CNR), Monserrato, Cagliari, Italy.
189. Department of Biomedical Sciences, University of Sassari, Sassari, Italy.
190. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.
191. Laboratory of Cardiovascular Science, NIA/NIH , Baltimore, MD 21224,USA.
192. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands.
193. Labormedizinisches Zentrum Dr. Risch, Landstrasse 157, 9494 Schaan, Liechtenstein.
194. Private University of the Principality of Liechtenstin, Dorfstrasse 24, 9495 Triesen, Liechtenstein.

195. University Insitute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland.
196. Department of Cardiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, PO Box 30001, 9700 RB Groningen, The Netherlands.
197. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA.
198. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA.
199. Nashville VA (TVHS) & Vanderbilt University.
200. VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School.
201. Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA, USA.
202. Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA.
203. National Institute for Health Research Imperial Biomedical Research Centre, Imperial College Healthcare NHS Trust and Imperial College London, London, UK.
204. UK Dementia Research Institute (UK DRI) at Imperial College London, London, UK

Corresponding authors: Mark Caulfield (m.j.caulfield@qmul.ac.uk) and Pau Elliott (p.elliott@imperial.ac.uk)

Abstract

High blood pressure is a highly heritable and modifiable risk factor for cardiovascular disease. We report the largest genetic association study of blood pressure traits (systolic, diastolic, pulse pressure) to date in over one million people of European ancestry. We identify 535 novel blood pressure loci that not only offer new biological insights into blood pressure regulation but also reveal shared genetic architecture between blood pressure and lifestyle exposures. Our findings identify new biological pathways for blood pressure regulation with potential for improved cardiovascular disease prevention in the future.

High blood pressure (BP) is a leading heritable risk factor for stroke and coronary artery disease and was responsible for an estimated 7.8 million deaths and 148 million disability life years lost worldwide in 2015 alone[1]. Studies indicate that an individual's blood pressure (BP) level is determined by complex interactions between life course exposures and their genetic background[2-4]. Previous genetic association studies have included genome-wide meta-analyses, customised cardiovascular candidate gene centric analyses and evaluation of exome variation. These have identified and validated variants at 274 loci, with modest effects on population BP that in aggregate explain only ~3% of the trait variance[5-12].

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] and the International Consortium of Blood Pressure-Genome Wide Association Studies (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 (MAF) $\geq$ 1% in association with BP traits (**Fig. 1**). We studied over 1 million people of European descent across both discovery and replication, including replication data from the US Million Veterans Program (MVP)[14] and the Estonian Genome Centre, University of Tartu (EGCUT) Biobank[15].

Briefly, UKB is a prospective cohort study of ~500,000 individuals recruited at ages 40-69 years who have been richly phenotyped including BP measurements[14]. Participants were genotyped using a customized array with imputation from the Haplotype Reference Consortium (HRC) panel, yielding ~7 million SNPs (imputation quality score (INFO) $\geq$ 0.1 and MAF $\geq$ 1%)[16]. After quality control (QC) and exclusions (Online Methods) we performed genome-wide association studies (GWAS) of BP traits using data from 458,577 UKB participants of European descent under an additive genetic model[17] (**Supplementary Table 1a**). Following LD-score regression[18], genomic control was applied to the UKB data prior to meta-analysis (Online methods).

In addition, we performed GWAS analyses for BP traits in the newly extended ICBP GWAS data comprising 77 independent studies including up to 299,024 participants of European ancestry genotyped with various arrays, and imputed to either the 1,000 Genomes Reference Panel or the HRC platforms (**Supplementary Table 1b**). After QC we applied genomic control at the individual study level and obtained summary effect sizes for ~7 million SNPs with INFO $\geq$ 0.3 and Cochran's Q statistic[19] (test of heterogeneity) filtered at $P \geq 1 \times 10^{-4}$ (Online Methods).

We then combined the UKB and ICBP GWAS results using inverse-variance weighted fixed effects meta-analysis (Online Methods), giving a total discovery sample of 757,601 individuals[20].

In our two-stage design we attempted replication 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) ($r^2 \geq 0.1$) or $\pm$500 Kb from any previously validated BP-associated SNPs at the 274 published loci. We used MVP (up to 220,520 people of European descent) and EGCUT (up to 28,742 Europeans) for independent external

replication[14,15] (**Supplementary Table 1c**). Our replication criteria for the two-stage design were genome-wide significance ($P < 5 \times 10^{-8}$) in the combined meta-analysis, with $P < 0.01$ in the replication data and concordant direction of effect between discovery and replication.

Given the larger size of the two discovery datasets (UKB and ICBP) compared with replication resources, we additionally undertook a one-stage design with internal replication, to minimize the risk of missing true positive associations from our two-stage analysis. To ensure the robustness of this approach, and to avoid false positive findings, we used $P < 5 \times 10^{-9}$ as the *P*-value threshold from the discovery meta-analysis, i.e. an order of magnitude more stringent than genome-wide significance[21]. We also required an internal replication *P*-value of < 0.01 in each of the UKB and ICBP GWAS analyses and with concordant direction of effect, to ensure support from both data sources.

We then explored the putative function of the BP associated signals using a range of *in silico* resources, including expression quantitative trait loci (eQTLs), tissue and DNase I site enrichment, long range chromatin interactions (Hi-C), pathway analysis and 'druggability'. We investigated metabolomic signatures associated with our novel sentinel SNPs, evaluated the overlap with lifestyle exposures that influence BP, and examined the co-occurrence of BP-associated loci with other complex traits and diseases. We also carried out conditional analyses using genome-wide complex trait analysis (GCTA)[22]. Finally, we developed a genetic risk score and performed analysis to model the impact that all BP-associated variants have on BP level, risk of hypertension (HTN), other cardiovascular diseases and on BP in non-European ancestries.

RESULTS

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**). Of the 325 two-stage variants, 204 would also have met the one-stage criteria, while 121 were uniquely identified from our two-stage design (**Fig. 3a**). Thus using this dual approach, we were able to identify large numbers of additional loci that would not have been detected by either the one- or two-stage designs alone, as well as finding considerable overlap (**Fig. 3a**). For SBP, the distributions of effect sizes of the one-stage loci (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) are similar within the discovery data ($P = 0.447$) (**Supplementary Fig. 2**). Of the 210 loci found only in the one-stage analysis of UKB and ICBP, 186 are also genome-wide significant ($P < 5 \times 10^{-8}$) in the combined meta-analysis of all four discovery and replication resources, with all variants, except one, having concordant direction of effect between discovery and replication (**Supplementary Tables 3a-c**). Of the remaining 24 SNPs which are not genome-wide significant overall in the combined meta-analysis, 10 still have concordant direction of effect.

We confirm previous findings with support in our data for all 274 published BP loci (**Supplementary Fig. 1 & 2** and **Supplementary Table 4**) and >95% of the previously reported SNPs covered within our data are genome-wide significant. Only 6 available SNPs

did not reach Bonferroni-significance, likely because they were originally identified from non-European ancestries (e.g. rs6749447, rs10474346, rs11564022), or from a gene-age interaction analysis (rs16833934). In addition, we confirmed a further 92 loci that had previously been reported but not replicated (**Supplementary Table 5**)[9]. Overall, with 274 previously reported loci confirmed, 92 loci replicated for the first time, and 535 novel loci identified here, there are 901 BP-associated loci in total.

Discovery of novel genetic loci for blood pressure

Of the 535 independent novel loci, 363 SNPs were associated with only one trait; 160 with two traits and 12 with all three BP traits (**Fig. 3b**), reflecting the inter-correlations between BP traits despite their different physiology.

From the conditional analyses using GCTA we additionally identified 163 independent secondary signals with MAF ≥ 1%, which are associated with BP at genome-wide significance level **(Supplementary Table 6)**. Of these 163 secondary signals, 19 SNPs are in LD ($r^2 \geq 0.1$) with previously reported secondary signals, resulting in a total of 144 new secondary signals identified here. Hence in total there are now over 1,000 independent BP signals reported.

The estimated SNP-wide heritability (h^2) of BP traits in our data was 0.213, 0.212 and 0.194 for SBP, DBP and PP respectively, and there is a gain in the percentage of BP variance explained. For example, for SBP, the percentage variance explained increased from 2.8 % for the 274 previously published loci to 5.7% for all sentinel and secondary SNPs identified at all 901 loci **(Supplementary Table 7)**.

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 $r^2 \geq 0.8$) to the nearest gene within 5kb of a SNP. There were 1644 genes in the novel loci and 962 genes in the known loci. Then we 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 gland[24] to identify distal associated genes. There were 498 novel loci that 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 long regulatory loop with the SNPs in the *SLC30A10* locus, and the *TGFBR1* promoter forms a 100kb loop with the *COL15A1* locus **(Supplementary Table 8)**.

Our eQTL analysis identified 60 novel loci with eQTLs in arterial tissue and 20 in adrenal tissue **(Supplementary Table 9);** this is a substantial increase over those identified in our previously published GWAS on ~140K UKB individuals[10]. An example is SNP rs31120122 which defines an aortic eQTL that affects expression of the *MED8* gene within the *SZT2*

locus. In combination with Hi-C interaction data in MSC this finding supports a role for *MED8* in BP regulation, possibly 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, FLRT2, ACVR2A*, *POU4F1*).

We investigated which transcription factors and chromatin marks are involved in regulatory interactions using the functional predictions from 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)**.

We used our genome-wide data at a false discovery rate (FDR) < 1% to robustly assess the 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. Due to the addition of our novel BP loci, we identified new findings from both the gene ontology and protein-protein interaction subnetwork enrichments, which highlight the TGFβ ($P = 2.3 \times 10^{-13}$) and related SMAD pathways ($P = 7 \times 10^{-15}$) (**Supplementary Table 10b, Supplementary Fig. 5b-d**).

We used FORGE[25] to investigate the regulatory regions for cell type specificity from DNase I hypersensitivity sites, which showed strongest enrichment ($P < 0.001$) in the vasculature and highly vascularised tissues, as reported in previous BP genetic studies[10] **(Supplementary Fig. 6).**

Potential therapeutic targets

Ingenuity pathway analysis and upstream regulator assessment showed enrichment of canonical pathways implicated in cardiovascular disease including pathways targeted by antihypertensive drugs (e.g. nitric oxide signalling) and also suggested some potential new targets, such as relaxin signalling. Notably, upstream regulator analysis identified several known mediators of BP including therapeutic targets such as angiotensinogen, calcium channels, progesterone, natriuretic 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.

We present a summary of the vascular expressed genes contained within the 535 novel loci, including a review of their potential druggability **(Supplementary Fig. 8)**. The overlap between genes associated with BP and those associated with antihypertensive drug targets, further demonstrates new genetic support for known drug mechanisms. For example, we

report five novel BP associations with the targets of five antihypertensive drug classes **(Supplementary Table 11).** These include the *PKD2L1, SLC12A2, CACNA1C, CACNB4 and CA7* loci, which are targeted by potassium-sparing diuretics (amiloride), loop diuretics (bumetanide and furosemide), dihydropyridine, calcium channel blockers, non-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

UK Biobank has collected extensive lifestyle related data, some of which are associated with BP epidemiologically and in trials. These include macronutrients, water, tea, caffeine and alcohol intake, anthropomorphic traits, physical activity and inactivity, smoking and urinary sodium, potassium and creatinine excretion[14]. We investigated whether sentinel SNPs at the 901 BP loci were associated with lifestyle traits in UKB in either the Stanford Global Biobank Engine (N = 327,302) or Gene ATLAS (N = 408,455), with corrected $P < 1 \times 10^{-6}$. For example, we found that a BP SNP rs34783010 in *GIPR* is associated with daily fruit intake ($P = 1.03 \times 10^{-7}$), urinary sodium and creatinine concentration ($P = 1.5 \times 10^{-13}$ and 1.2×10^{-9} respectively), body mass index (BMI, $P = 3.3 \times 10^{-41}$), weight ($P = 7.3 \times 10^{-35}$) and waist circumference ($P = 7.7 \times 10^{-30}$); rs6495122, near *CPLX3* and *ULK3*, and rs1378942 in *CSK* are associated with water ($P = 1.3 \times 10^{-22}$ and 2.6×10^{-20} respectively), caffeine ($P = 1.3 \times 10^{-46}$ and 2.2×10^{-43}) and tea intake ($P = 7.6 \times 10^{-38}$ and 8.1×10^{-33}), as well as urinary creatinine concentrations ($P = 5.6 \times 10^{-8}$ and $P = 3.2 \times 10^{-8}$ respectively). In addition, the BP 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 ($P = 4.7 \times 10^{-11}$) **(Supplementary Table 12).** We used unsupervised hierarchical clustering for the 36 BP loci that showed at least one association with the lifestyle related traits in UKB at $P < 1 \times 10^{-6}$ (**Fig. 4).** The heatmap summarises the locus specific associations across the range of traits and highlights heterogeneous effects with anthropometric traits across the range of loci examined. For example, it shows a cluster of associations between BP raising alleles and increased adult height and weight and another cluster of genes that show associations between BP raising alleles and decreased adult height and weight. We note that some observed cross-trait associations are in counter-directions to what may be expected epidemiologically.

Association lookups with other traits and diseases

We further evaluated cross-trait and disease associations using GWAS catalog[26], PhenoScanner[27] and DIsGeNET, which integrates data from expert curated repositories, GWAS catalogues, animal models and the literature[28,29]. The GWAS catalog and PhenoScanner search of published GWAS showed that 77 of our 535 novel loci (using sentinel SNPs or proxies; $r^2 \geq 0.8$) are also significantly associated with other traits and diseases (**Fig. 5, Supplementary Table 13**). We identified *APOE* as a highly cross-related BP locus showing associations with lipid levels, cardiovascular related outcomes and Alzheimer's disease, a finding that highlights a common link between cardiovascular risk and

cognitive decline (**Fig. 5**). Several other loci overlap with anthropometric traits, including BMI, birth weight and height (**Fig 5)**. DisGeNET terms related to lipid measurements, cardiovascular outcomes and obesity overlap with BP loci **(Fig. 6)**.

We used ^{1}H NMR lipidomics data on plasma (N=2,022) and data from the Metabolon platform (N=1,941) for subsets of participants of the Airwave Health Monitoring Study[30] for lookups of our sentinel SNPs. We also used PhenoScanner to test each SNP against published significant ($P < 5 \times 10^{-8}$) genome vs metabolome-wide associations in plasma and urine (Online Methods). Ten BP SNPs show association with lipid particle metabolites and a further 31 SNPs (8 also on PhenoScanner) show association with metabolites on the Metabolon platform, highlighting lipid pathways, amino acids (glycine, serine, glutamine), tri-carboxylic acid cycle intermediates (succinylcarnitine) and drug metabolites (**Supplementary Tables 14 and 15**). These findings suggest a close metabolic coupling of BP regulation with lipid, and for the first time, with energy metabolism.

Genetic risk of increased blood pressure, hypertension and cardiovascular disease

We created a genetic risk score (GRS) for BP levels weighted according to the effect estimates from ICBP (for known loci) and the MVP+EGCUT replication (for novel loci) across all 901 loci (Online Methods). The combination of these BP variants was associated with a 10.39 mmHg higher, sex-adjusted mean SBP in UK Biobank for the comparison between the upper and lower quintiles of the GRS distribution (95% CI: 10.19 to 10.58 mm Hg, $P < 1 \times 10^{-300}$) and with 12.85 mmHg difference in SBP (95% CI: 12.57 to 13.13, $P < 1 \times 10^{-300}$) comparing the upper and lower deciles (**Fig. 7a, Supplementary Table 16**). In addition, we observed over two-fold sex-adjusted higher risk of hypertension (OR 2.66; 95% CI: 2.60 to 2.72; $P < 1 \times 10^{-300}$) between the upper and lower quintiles of the GRS in UK Biobank (**Fig. 7**). Sensitivity analyses in the independent Airwave cohort gave similar results (**Supplementary Table 17**).

From record linkage to Hospital Episode Statistics and mortality follow-up in UKB we show that the GRS is associated with increased, sex-adjusted risk of incident stroke, myocardial infarction and all incident cardiovascular outcomes, comparing the upper and lower deciles of the GRS distribution, with odds ratios of 1.47 (95% CI: 1.35 to 1.59, $P = 1.12 \times 10^{-20}$), 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

We examined associations with BP of both individual SNPs and the GRS among unrelated individuals of African and of South Asian ancestries in UKB, for the 901 known and novel loci. In comparison with the results for European ancestry, 62.4%, 62.5% and 64.8% of the variants among people of African (N=7,782), and 74.2%, 72.3% and 75% of South Asian (N=10,323) ancestry 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 r^2= 0.37 and 0.78 for African and South Asian ancestries respectively **(Supplementary Fig. 10)**. We then applied the GRS derived from European ancestry findings to unrelated individuals of African (N=6,970) and South Asian (N=8,827)

ancestries. BP variants in combination were associated with 6.1 mmHg (95% CI: 4.50 to 7.65; $P = 4.9 \times 10^{-14}$) and 7.4 mmHg (95% CI: 6.00 to 8.69; $P = 1.7 \times 10^{-26}$) higher, sex-adjusted mean systolic pressure among individuals of African and South Asian ancestries, respectively, for the comparison between the upper and lower quintiles of the GRS distribution **(Supplementary Tables 19a and 19b)**.

DISCUSSION

Our study of over 1 million people offers an important step forward in understanding the genetic architecture of BP. We have identified over 1,000 independent signals at 901 loci for BP traits, and with 535 novel loci we have more than tripled the number of BP loci and doubled the percentage variance explained for BP. By now explaining 27% of the estimated heritability for BP, we make major inroads into the missing heritability influencing BP level in the population[31]. These findings illustrate the power of a large-scale standardised approach to data collection, biobanking, genotyping, quality control and imputation, such as was achieved in UKB. The novel loci open the vista of entirely new biology and highlight gene regions in systems not previously implicated in BP regulation. This is particularly timely as the global prevalence of people with SBP over 110-115 mm Hg, above which cardiovascular risk increases in a continuous graded manner, now exceeds 3.5 billion and those within the treatment range exceed 1 billion[32,33].

Our functional analysis highlights the role of the vasculature and associated pathways in the genetics underpinning BP traits. We show a role for several loci in the transforming growth factor beta (TGFβ) pathway including SMAD family genes and the *TGFβ* gene locus itself. This pathway affects sodium handling in the kidney, ventricular remodelling and recently plasma levels of TGFβ have been correlated with hypertension **(Fig. 8)**[34,35]. The activin A receptor type 1C (*ACVR1C*) gene mediates the effects of the TGFβ family of signalling molecules. Another BP locus contains the Bone Morphogenetic Protein 2 (*BMP2*) gene in the TGFβ pathway, which prevents growth suppression in pulmonary arterial smooth muscle cells and is associated with pulmonary hypertension[36]. We identified another BP locus including the Kruppel-like family 14 (*KLF14*) gene of transcription factors which are induced by low levels of TGFβ receptor II gene expression. This gene has also been associated with type 2 diabetes, hypercholesterolaemia and atherosclerosis[37].

Our analysis shows enrichment of BP genes in the adrenal tissue. The adrenal gland has a key role in BP regulation, with autonomous aldosterone production by the adrenal glands thought to be responsible for 5-10% of all hypertension, rising to ~20% amongst people with resistant hypertension[38]. Some of our novel loci are linked functionally to aldosterone secretion[39,40]. For example, the *CTNNB1* locus encodes β-catenin, the central molecule in the canonical Wnt signalling system, required for normal adrenocortical development[41,42]. Somatic adrenal mutations of this gene that prevent serine/threonine phosphorylation lead to hypertension through the generation of aldosterone-producing adenomas[43,44].

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 enalapril[45]. We previously reported a fibroblast growth factor (*FGF5*) gene locus in association with BP. Here, we additionally identify a new BP locus encoding FGF9, which has been linked to enhanced angiogenesis and vascular smooth muscle cell differentiation by regulating *VEGFA* expression.

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

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 of the novel loci for SBP contains the relaxin gene which encodes a G-protein coupled receptor with roles in vasorelaxation and cardiac function, and which signals by phosphatidylinositol 3-kinase (PI3K)[48,49], an enzyme which inhibits vascular smooth muscle cell proliferation and neo-intimal formation[50]. We identify the *PI3K* gene here as a novel BP locus. We also identify the novel *RAMP2* locus which encodes an adrenomedullin receptor[51]; we previously identified the adrenomedullin (*ADM*) gene as a BP locus[12]. 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 which interacts with Rho GTPases involved in contraction of vascular smooth muscle and regulation of responses to angiotensin II[52].

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

Our findings with larger sample size, strengthen our previously reported genetic risk score 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, giving substantially increased risk of cardiovascular events (stroke and coronary disease)[10]. 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 ages[10].

We identified several BP-associated loci which are also associated with lifestyle traits, suggesting a possible shared genetic architecture between BP and lifestyle exposures known from randomised clinical trials and observational data to be associated with BP[53]. We adjusted our BP GWAS analyses for BMI, which should have reduced possible confounding effects, though we acknowledge the potential for collider bias between a lifestyle factor and BP[54]. Nonetheless, our findings of possible genetic overlap between loci associated with BP and lifestyle exposures could support renewed focus on altering specific lifestyle measures known to affect BP[55].

Despite smaller sample sizes in the single-variant analyses of non-European ancestry samples, we observed high concordance of the effects of BP variants in people of African (> 62%) and South Asian (> 72%) ancestry for all BP traits. Furthermore, the GRS analyses show that, in combination, BP variants identified in European analyses are also associated with BP in non-European ancestries, though size of the effect was 30-40% smaller. Nonetheless, new knowledge on the genetic architecture of BP in Europeans also extends, at least in part, to populations of other ancestries.

Our discovery of 535 novel loci from a combination of a two-stage GWAS with independent replication and one-stage GWAS design with internal replication illustrates the value of this approach to minimize the effects of stochastic variation and heterogeneity. The one-stage approach was included in order to report signals that had independent and concordant support ($P < 0.01$) from both UKB and ICBP (thus minimizing the impact of winners' curse on our reported findings). Indeed, all but two of the 210 SNPs discovered in the one-stage analysis reach $P < 5 \times 10^{-6}$ in either UKB or ICBP (the two exceptions have $P = 7.8 \times 10^{-6}$ in UKB, $P = 1.4 \times 10^{-5}$ in ICBP; and $P = 1.2 \times 10^{-5}$ in UKB, $P = 5.3 \times 10^{-6}$ in ICBP). To further minimize the risk of reporting false positive loci within our one-stage design, we set a stringent overall discovery meta-analysis *P*-value threshold of $P < 5 \times 10^{-9}$, an order of magnitude smaller than a genome-wide significance *P*-value, in line with thresholds recommended for whole genome sequencing[22]. We also found high concordance in the direction of effects between discovery data in the one-stage approach and the replication resources, with similar distributions of effect sizes for the one-stage vs two-stage discovery data. We note that 24 of the one-stage SNPs which reached $P < 5 \times 10^{-9}$ in discovery failed to reach genome-wide 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 studies if larger replication resources became available.

The new discoveries reported here more than triple the number of loci for BP to a total of 901 and represent a substantial advance in understanding the genetic architecture of BP. By identifying so many novel genes across the genome, these findings could partly support an omnigenic model for complex traits where genome-wide association of multiple interconnected pathways is observed. However, as our strong tissue enrichment shows particular relevance to the biology of BP and cardiovascular disease[56], the genetic architecture of BP still shows 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.

URLs

FORGE: http://browser.1000genomes.org/Homo_sapiens/UserData/Forge?db=core
Fantom5 data: http://fantom.gsc.riken.jp/5/
ENCODE DNase I data: (wgEncodeAwgDnaseMasterSites; accessed using Table browser)
ENCODE cell type data: http://genome.ucsc.edu/ENCODE/cellTypes.html.
GTEx: www.gtexportal.org
DeepSEA: http://deepsea.princeton.edu/
WebGetstalt: http://www.webgestalt.org
IPA: www.qiagen.com/ingenuity
Mouse Genome Informatics (MGI): http://www.informatics.jax.org/batch
Drug Gene Interaction database: www.dgidb.org
PhenoScanner: http://www.phenoscanner.medschl.cam.ac.uk (Phenoscanner integrates results from the GWAS catalogue: https://www.ebi.ac.uk/gwas/ and GRASP: https://grasp.nhlbi.nih.gov/)
DisGeNEt: http://www.disgenet.org
GeneATLAS: http//geneatlas.roslin.ed.ac.uk
Global Biobank Engine: https://biobankengine.stanford.edu

References

1. Forouzanfar, M.H. *et al.* Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. *JAMA* **317**, 165-182 (2017).
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).
3. Poulter, N.R., Prabhakaran, D. & Caulfield, M. Hypertension. *Lancet* **386**, 801-12 (2015).
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).
5. 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).
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).
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).
8. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
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).

10. Warren, H.R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* **49**, 403-415 (2017).
11. Wain, L.V. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* (2017).
12. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
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).
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).
15. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* **44**, 1137-47 (2015).
16. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
17. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
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).
19. Ioannidis, J.P., Patsopoulos, N.A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* **2**, e841 (2007).
20. Evangelou, E. & Ioannidis, J.P. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* **14**, 379-89 (2013).
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).
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).
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).
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).
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. *F1000Research* **4**(2015).
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).
27. Staley, J.R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207-3209 (2016).
28. Pinero, J. *et al.* DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* **2015**, bav028 (2015).
29. Pinero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).

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).
31. Ehret, G.B. & Caulfield, M.J. Genes for blood pressure: an opportunity to understand hypertension. *Eur Heart J* **34**, 951-61 (2013).
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 assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1659-1724 (2016).
34. 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).
35. Feng, W., Dell'Italia, L.J. & Sanders, P.W. Novel Paradigms of Salt and Hypertension. *J Am Soc Nephrol* **28**, 1362-1369 (2017).
36. International PPH Consortium *et al.* Heterozygous germline mutations in BMPR2, encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. *Nat Genet* **26**, 81-4 (2000).
37. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* **42**, 579-89 (2010).
38. Douma, S. *et al.* Prevalence of primary hyperaldosteronism in resistant hypertension: a retrospective observational study. *Lancet* **371**, 1921-6 (2008).
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).
40. Calhoun, D.A., Nishizaka, M.K., Zaman, M.A., Thakkar, R.B. & Weissmann, P. Hyperaldosteronism among black and white subjects with resistant hypertension. *Hypertension* **40**, 892-6 (2002).
41. Drelon, C., Berthon, A., Mathieu, M., Martinez, A. & Val, P. Adrenal cortex tissue homeostasis and zonation: A WNT perspective. *Mol Cell Endocrinol* **408**, 156-64 (2015).
42. El Wakil, A. & Lalli, E. The Wnt/beta-catenin pathway in adrenocortical development and cancer. *Mol Cell Endocrinol* **332**, 32-7 (2011).
43. Teo, A.E. *et al.* Pregnancy, Primary Aldosteronism, and Adrenal CTNNB1 Mutations. *N Engl J Med* **373**, 1429-36 (2015).
44. Tissier, F. *et al.* Mutations of beta-catenin in adrenocortical tumors: activation of the Wnt signaling pathway is a frequent event in both benign and malignant adrenocortical tumors. *Cancer Res* **65**, 7622-7 (2005).
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).
46. Yang, R. *et al.* Hypertension and endothelial dysfunction in apolipoprotein E knockout mice. *Arterioscler Thromb Vasc Biol* **19**, 2762-8 (1999).
47. Sofat, R. *et al.* Circulating Apolipoprotein E Concentration and Cardiovascular Disease Risk: Meta-analysis of Results from Three Studies. *PLoS Med* **13**, e1002146 (2016).
48. Conrad, K.P. Unveiling the vasodilatory actions and mechanisms of relaxin. *Hypertension* **56**, 2-9 (2010).

49. Sun, H.J. *et al.* Relaxin in paraventricular nucleus contributes to sympathetic overdrive and hypertension via PI3K-Akt pathway. *Neuropharmacology* **103**, 247-56 (2016).
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).
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).
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. *Arterioscler Thromb Vasc Biol* **25**, 1831-6 (2005).
53. 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).
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).
55. Pazoki, R. et al. Genetic predisposition to high blood pressure and lifestyle factors: Associations with midlife blood pressure levels and cardiovascular events. *Circulation* **137,** 653-661 (2018)
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)

Acknowledgements

H.R.W. was funded by the National Institute for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Centre at Barts and The London School of Medicine and Dentistry. D.M-A is supported by the Medical Research Council [grant number MR/L01632X.1]. B.M. holds an MRC eMedLab Medical Bioinformatics Career Development Fellowship, funded from award MR/L016311/1. H.G. was funded by the NIHR Imperial College Health Care NHS Trust and Imperial College London Biomedical Research Centre. C.P.C. was funded by the National Institute for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Center at Barts and The London School of Medicine and Dentistry. S.T. was supported by Canadian Institutes of Health Research; Université Laval (Quebec City, Canada). G.P. was supported by Canada Research Chair in Genetic and Molecular Epidemiology and CISCO Professorship in Integrated Health Biosystems. I.K. was supported by the EU PhenoMeNal project (Horizon 2020, 654241). C.P.K. is supported by grants R01DK096920 and U01DK102163 from the NIH-NIDDK, and by resources from the Memphis VA Medical Center. P.S.T. was supported by VA Award #I0BX003362. C.P.K. is an employee of the US Department of Veterans affairs. Opinions expressed in this paper are those of the authors' and do not necessarily represent the opinion of the Department of Veterans Affairs. S.D. was supported for this work by grants from the European Research Council (ERC), the EU Joint Programme - Neurodegenerative Disease Research (JPND), the Agence Nationale de la Recherche (ANR). M.BOE is supported by NIH grant R01-DK062370. H.W. and A.G. acknowledge support of the Tripartite Immunometabolism Consortium [TrIC], Novo Nordisk Foundation (grant

NNF15CC0018486). N.V. was supported by Marie Sklodowska-Curie GF grant (661395) and ICIN-NHI. C.M. is funded by the MRC AimHy (MR/M016560/1) project grant. M.A.N participation is supported by a consulting contract between Data Tecnica International and the National Institute on Aging, NIH, Bethesda, MD, USA. M.BR., M.CO., I.G., P.G., G.G, A.MO., A.R., D.V. were supported by the Italian Ministry of Health RF2010 to Paolo Gasparini, RC2008 to Paolo Gasparini. D.I.B. is supported by the Royal Netherlands Academy of Science Professor Award (PAH/6635). J.C.C. is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award (NMRC/STaR/0028/2017). C.C., P.B.M and M.R.B were funded by the National Institutes for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Centre at Barts. M.R. is recipient from China Scholarship Council (No. 2011632047). C.L. was supported by the Medical Research Council UK (G1000143; MC_UU_12015/1; MC_PC_13048; MC_U106179471), Cancer Research UK (C864/A14136), EU FP6 programme (LSHM_CT_2006_037197). G.B.E is supported by the Swiss National Foundation SPUM project FN 33CM30-124087, Geneva University, and the Fondation pour Recherches Médicales, Genève. C.M.L is supported by the Li Ka Shing Foundation, WT-SSI/John Fell funds and by the NIHR Biomedical Research Centre, Oxford, by Widenlife and NIH (CRR00070 CR00.01). R.J.F.L. is supported by the NIH (R01DK110113, U01HG007417, R01DK101855, R01DK107786). D.O.M-K. is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023). M.M was supported by the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. H.W. and M.F. acknowledge the support of the Wellcome Trust core award (090532/Z/09/Z) and the BHF Centre of Research Excellence (RE/13/1/30181). A.G, H.W. acknowledge European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no. HEALTH-F2-2013-601456 (CVGenes@Target) & and A.G, the Wellcome Trust Institutional strategic support fund. A.P.M. was a Wellcome Trust Senior Research Fellow in Basic Biomedical Science (WT098017). L.R. was supported by Forschungs- und Förder-Stiftung INOVA, Vaduz, Liechtenstein. M.TO. is supported by British Heart Foundation (PG/17/35/33001). P.v.d.H. was supported by ICIN-NHI and Marie Sklodowska-Curie GF (call: H2020-MSCA-IF-2014, Project ID: 661395). N.J.W. was supported by the Medical Research Council UK (G1000143; MC_UU_12015/1; MC_PC_13048; MC_U106179471), Cancer Research UK (C864/A14136), EU FP6 programme (LSHM_CT_2006_037197). E.Z. was supported by the Wellcome Trust (WT098051). J.N.H. and A.G. were supported by VA grant 1I01CX000982. T.L.E. and D.R.V.E. were supported by grant R21HL121429 from NIH/NHLBI. A.M.H. was supported by VA Award #I01BX003360. C.J.O. was supported by the VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School. The MRC/BHF Cardiovascular Epidemiology Unit is supported by the UK Medical Research Council [MR/L003120/1]; British Heart Foundation [RG/13/13/30194]; and UK National Institute for Health Research Cambridge Biomedical Research Centre. J.DA is a British Heart Foundation Professor and NIHR Senior Investigator. L.V.W. holds a GlaxoSmithKline/British Lung Foundation Chair in Respiratory Research. P.E. acknowledges support from the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre, the NIHR Health Protection Research Unit in Health Impact of

Environmental Hazards (HPRU-2012-10141), and the Medical Research Council (MRC) and Public Health England (PHE) Centre for Environment and Health (MR/L01341X/1). P.E. is a UK Dementia Research Institute (DRI) professor, UK DRI at Imperial College London, funded by the MRC, Alzheimer's Society and Alzheimer's Research UK. M.J.C. was funded by the National Institute for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Biomedical Research Center at Barts and The London School of Medicine and Dentistry. M.J.C. is a National Institute for Health Research (NIHR) senior investigator and this work is funded by the MRC eMedLab award to M.J.C. and M.R.B. and the NIHR Biomedical Research Centre at Barts.

This research was supported by the British Heart Foundation (grant SP/13/2/30111). Project title: Large-Scale Comprehensive Genotyping of UK Biobank for Cardiometabolic Traits and Diseases: UK CardioMetabolic Consortium (UKCMC). This research has been conducted using the UK Biobank Resource under application number 236.

Computing: This work was enabled using the computing resources of the i) UK MEDical BIOinformatics partnership - aggregation, integration, visualisation and analysis of large, 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].

Conflicts/Disclosures

M.A.N. consults for Illumina Inc, the Michael J. Fox Foundation and University of California Healthcare among others.M.J.C. is Chief Scientist for Genomics England, a UK Government company.

The views expressed in this manuscript are those of the authors and do not necessarily 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 represent the views of the Department of Veterans Affairs or the United States Government.

Author contributions

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., K.W., E.T. L.V.W.

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., M.J.C. (with group leads EE, H.R.W, L.V.W., P.E., M.J.C.)

ICBP-Discovery contributor: (3C-Dijon) S.D., M.S., P.AM., G.C., C.T.; (AGES-Reykjavik) V.GU., 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., 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., 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., 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., A.S.B., J.DA., J.M.M.H.; (EPIC-Norfolk, Fenland-OMICS, Fenland-GWAS) J-H.Z.; (EPIC-Norfolk,

Fenland-OMICS, Fenland-GWAS, InterAct-GWAS) J.L., C.L., R.A.S., N.J.W.; (ERF) N.A., B.A.O., C.M.v.D.; (Fenland-Exome, EPIC-Norfolk-Exome) S.M.W., FHSS-J.H., D.L.; (FINRISK (COROGENE_CTRL)) P.J., K.K., M.P., A-P.S.; (FINRISK_PREDICT_CVD) A.S.H., A.P., S.R., V.S.; (FUSION) A.U.J, M.BOE., F.C., J.T., (GAPP) S.T., G.P., D.CO., L.R.; (Generation Scotland (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., 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.; (HCS) 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., 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., P.G., D.V.; (INGI-VB) C.M.B., N.J.S., D.T., M.T.; (JUPITER) F.G., L.M.R., P.M.R., D.I.C.; (KORA 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.; (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) 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., 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., 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) 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., 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., 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., 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) M.M., C.M., T.D.S.; (UKHLS) B.P., E.Z.; (ULSAM) V.G., A.P.M., A.M., E.I.; (WGHS) F.G., L.M.R., P.M.R., D.I.C.; (YFS) M.K., T.L., L-P.L., O.T.R.

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., P.S.T., C.P.K., A.M.H., C.J.O., T.L.E.; (EGCUT) T.E., R.M., L.M. A.ME.

Airwave Health Monitoring Study: E.E, H.G, A-C.V., R.P., I.K., I.T., P.E.

All authors critically reviewed and approved the final version of the manuscript

Figure 1. Study design schematic for discovery and validation of loci. ICBP; International 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; linear mixed model; UKB, UK Biobank, MAF, minor allele frequency; HLA, Human Leukocyte Antigen; MVP, Million Veterans Program; EGCUT; Estonian Genome Center, University of Tartu; SBP, systolic blood pressure; DBP, diastolic blood pressure; PP, pulse pressure.

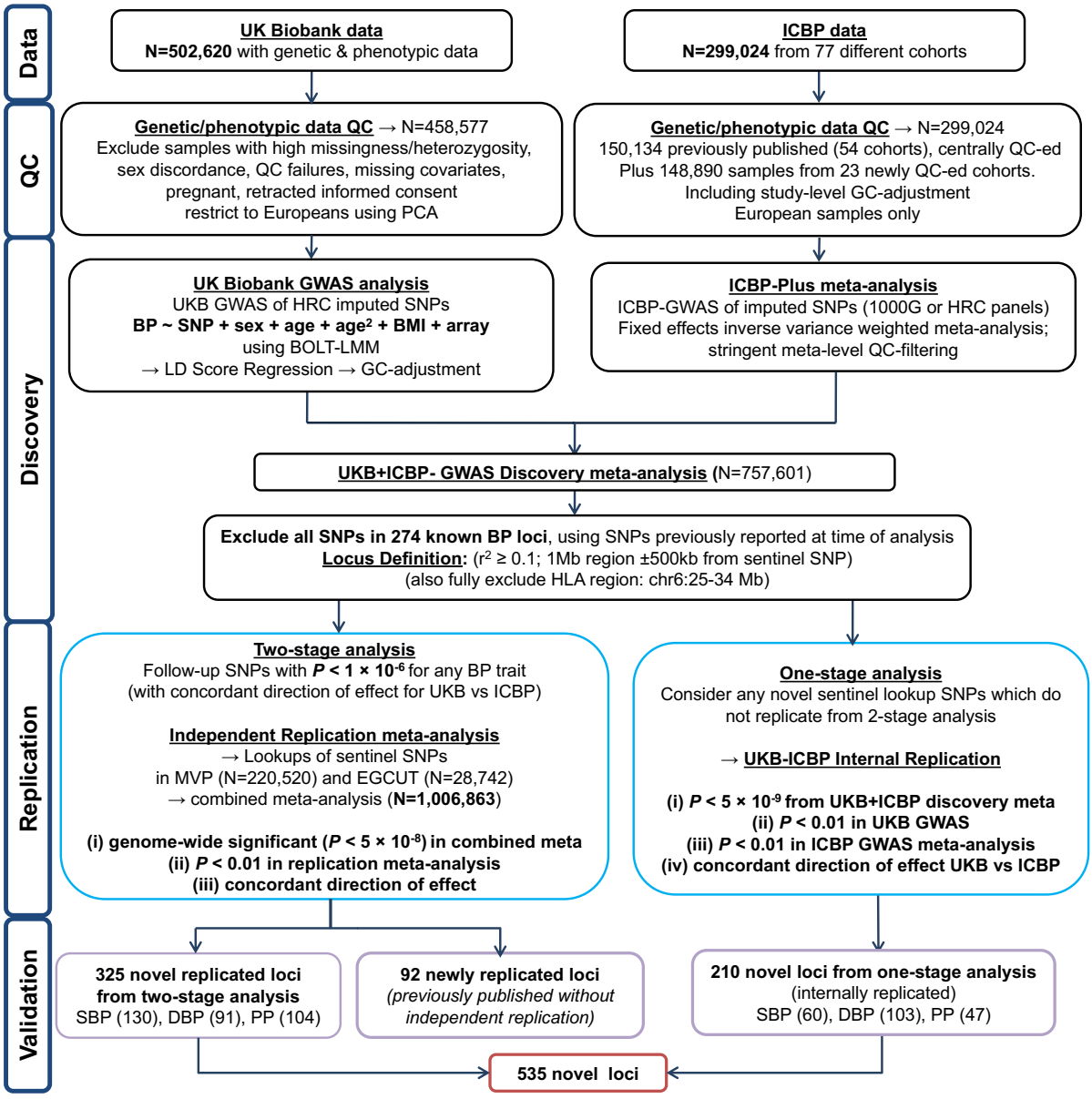


Figure 2. Manhattan plot showing the minimum *P*-value for the association across all blood pressure traits in the discovery stage excluding known and previously reported variants.
Manhattan plot of the discovery genome-wide association meta-analysis in 757,601 individuals excluding variants in 274 known loci. The minimum *P*-value across SBP, DBP and PP is presented. The y axis shows the $-\log_{10} P$ values and the x axis shows their chromosomal positions. Horizontal red and blue line represents the thresholds 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 from the 2-stage design whereas SNPs in red are in LD ($r^2 > 0.8$) with 210 SNPs identified through the 1-stage design with internal replication. Any loci in black or grey that exceed the significance thresholds were significant in the discovery meta-analysis, but did not meet the criteria of replication in the one- or two-stage designs.

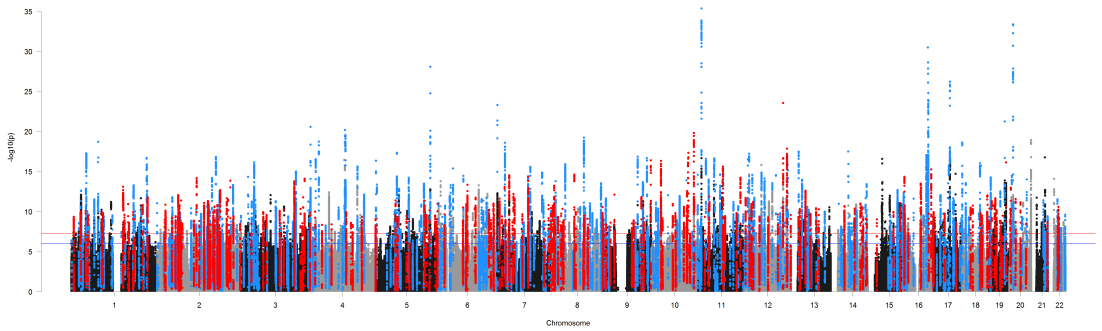


Figure 3: Venn Diagrams of Novel Loci Results (a) "Comparison of 1-stage and 2-stage design analysis criteria": For all 535 novel loci, we compare the results according to the association criteria used for the one-stage and the two-stage design. Two-hundred and ten loci exclusively met the one-stage analysis criteria (P <5x10^{-9} in the discovery meta-analysis, P < 0.01 in UKB, P < 0.01 in ICBP and concordant direction of effect between UKB and ICBP). Of the 325 novel replicated loci from the 2-stage analysis (genome-wide significance in the combined meta-analysis, P < 0.01 in the replication meta-analysis and concordant direction of effect), 204 loci would also have met the one-stage criteria, whereas 121 were only identified by the two-stage analysis. **(b) "Overlap of Associations across Blood Pressure Traits"**. For all 535 novel loci, we show the blood pressure traits associated with each locus. We present the two-stage loci first, followed by the one-stage loci. The locus names provided in alphabetical order correspond to the nearest annotated gene. SNPs: Single nucleotide polymorphisms; SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure; UKB: UK Biobank; ICBP: International Consortium of Blood Pressure

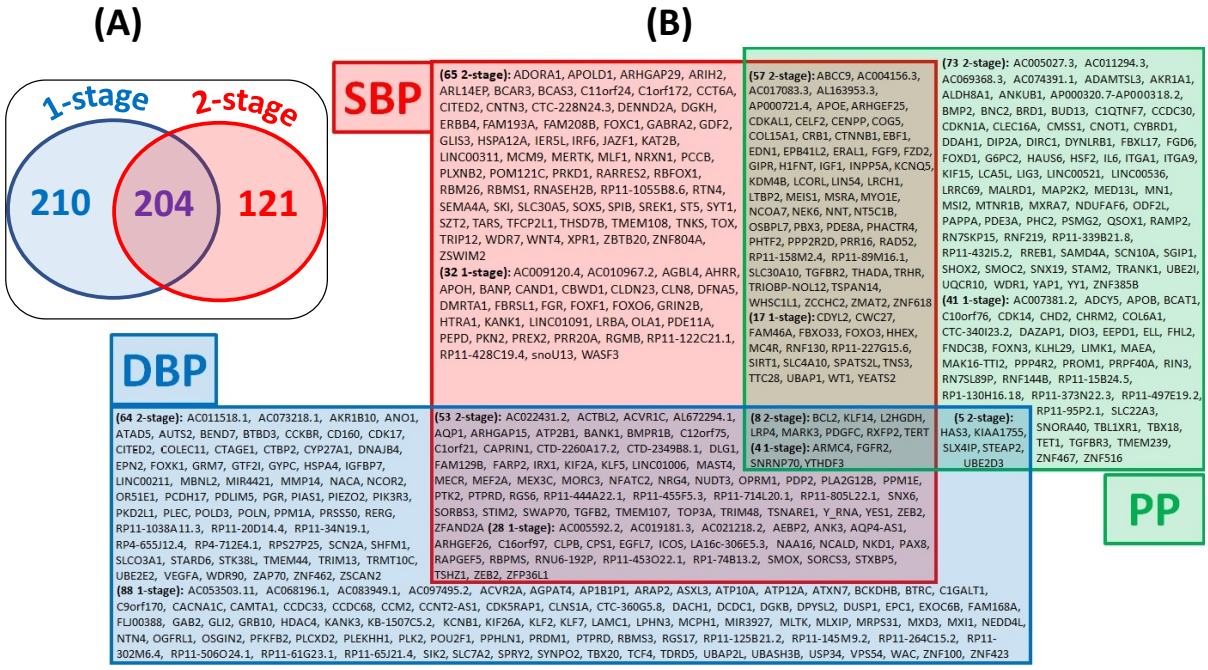


Figure 4. Association of blood pressure loci with lifestyle traits. Plot shows hierarchical clustering of BP loci based on associations with lifestyle-related factors. For the sentinel SNP at each BP locus (x axis), we calculated the -$\log_{10}(P)$*sign(β) (aligned to BP-raising allele) as retrieved from the Gene Atlas catalogue (http://geneatlas.roslin.ed.ac.uk). BP loci and traits were clustered according to the Euclidean distance amongst -$\log_{10}(P)$*sign(β). Red squares indicate direct associations with the trait of interest and blue squares inverse associations. Only SNPs with at least one association at $P < 10^{-6}$ with at least one of the traits examined are annotated in the heat-map. All 901 loci are considered, both known and novel: novel loci are printed in bold font. SNPs: Single Nucleotide Polymorphisms; BP: Blood Pressure.

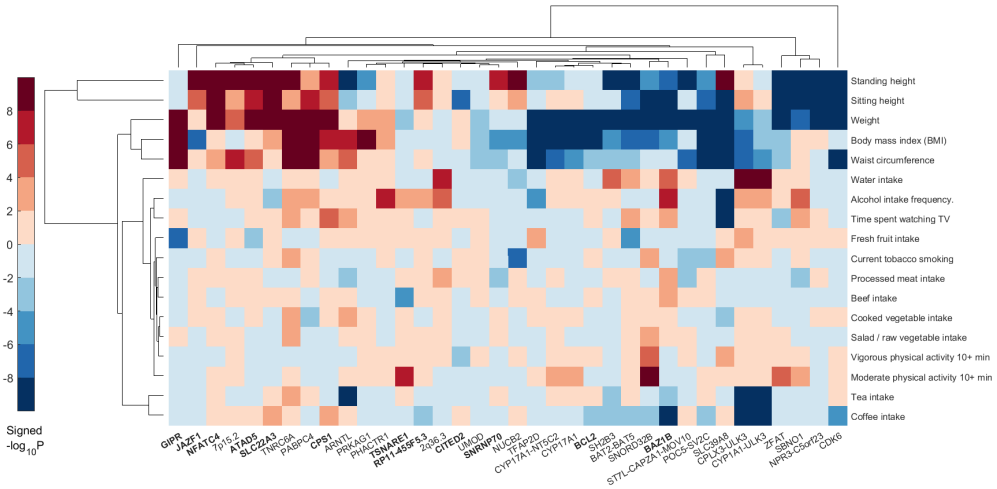
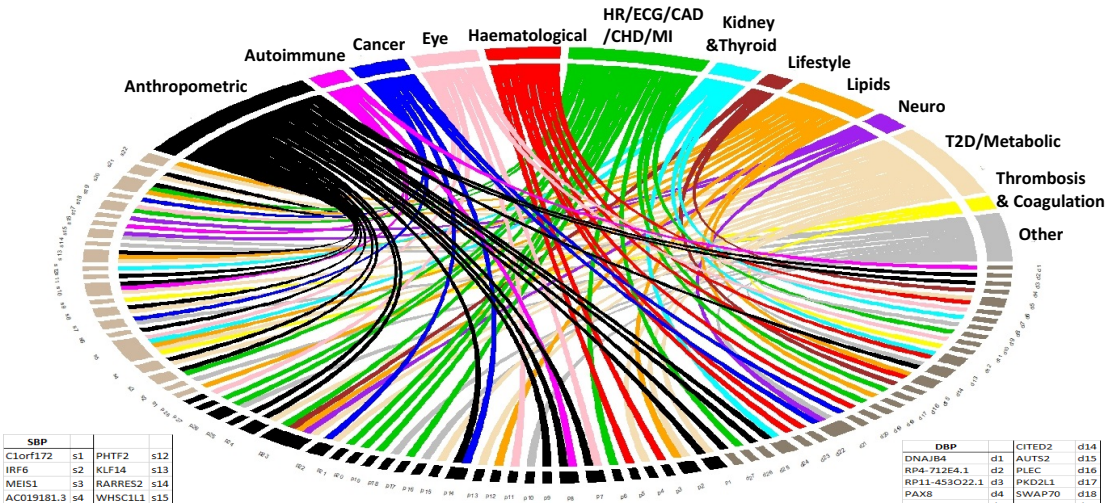


Figure 5. Association of blood pressure loci with other traits. Plot shows results from associations with other traits which were extracted from the GWAS catalog and PhenoScanner databases for the 535 novel sentinel SNPs including proxies in Linkage Disequilibrium ($r^2 \geq 0.8$) with genome-wide significant 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.



SBP			
C1orf172	s1	PHTF2	s12
IRF6	s2	KLF14	s13
MEIS1	s3	RARRES2	s14
AC019181.3	s4	WHSC1L1	s15
CPS1	s5	Y_RNA	s16
LCORL	s6	TSNARE1	s17
TERT	s7	HTRA1	s18
CDKAL1	s8	PPM1E	s19
STXBP5	s9	MC4R	s20
JAZF1	s10	BCL2	s21
POM121C	s11	PEPD	s22

PP									
SLC30A10	p1	FNDC3B	p7	NDUFAF6	p13	RP11-497E19.2	p19	BMP2	p25
APOB	p2	RN7SL89P	p8	COL15A1	p14	CNOT1	p20	KIAA1755	p26
AC007381.2	p3	FBXL17	p9	MTNR1B	p15	LIG3	p21	AP000320.7,AP000318.2	p27
G6PC2	p4	EBF1	p10	YAP1	p16	ELL	p22	TRIOBP,NOL12	p28
ZNF385B	p5	RREB1	p11	BCAT1	p17	APOE	p23		

DBP			
DNAJB4	d1	AUTS2	d15
RP4-712E4.1	d2	PLEC	d16
RP11-453O22.1	d3	PKD2L1	d17
PAX8	d4	SWAP70	d18
CCNT2-AS1	d5	FAM168A	d19
ZEB2	d6	POLD3	d20
ACVR2A	d7	GAB2	d21
HDAC4	d8	CACNA1C	d22
RP11-145M9.2	d9	NACA	d23
DUSP1	d10	MLXIP	d24
MXD3	d11	NCOR2	d25
VEGFA	d12	CTAGE1	d26
BCKDHB	d13	CDK5RAP1	d27
CITED2	d14		

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 r^2>0.8 with the known (red bars) SNPs from the 274 published loci, and all (green bars) BP variants from all 901 loci. Numbers on top of the bars denote the number of SNPs included in DisGeNET for the specific trait or disease. Traits/diseases with an overlap of at least 5 variants in LD with all markers are shown. The Y axis shows the percentage of variants associated with the diseases that is covered by the overlap. For the sake of clarity, the DisGeNET terms for blood pressure and hypertension are not displayed, whereas the following diseases have been combined: coronary artery disease (CAD), coronary heart disease (CHD) and myocardial infarction (MI); prostate and breast carcinoma; Crohn's and inflammatory bowel diseases.

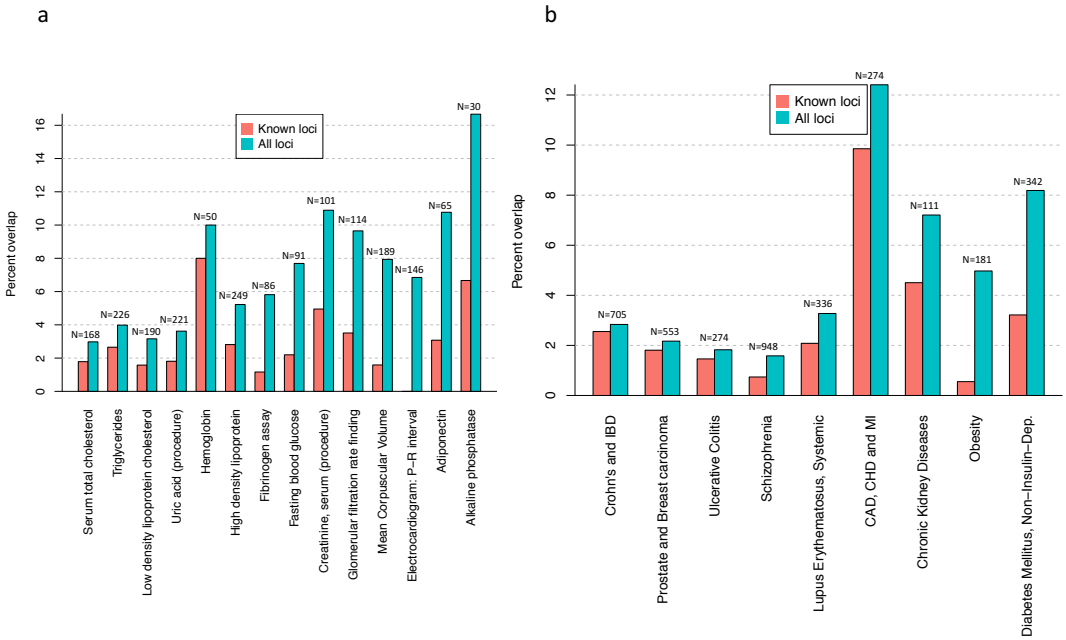


Figure 7. **Relationship of deciles of the genetic risk score (GRS) based on all 901 loci with blood pressure, risk of hypertension and cardiovascular disease in UK Biobank**. The plots show sex-adjusted (**a**) mean systolic blood pressure (SBP) and odds ratios of hypertension (HTN) and (**b**) odds ratios of incident cardiovascular disease (CVD), myocardial infarction (MI) and stroke, comparing each of the upper nine GRS deciles with the lowest decile; dotted lines represent the upper 95% confidence intervals.

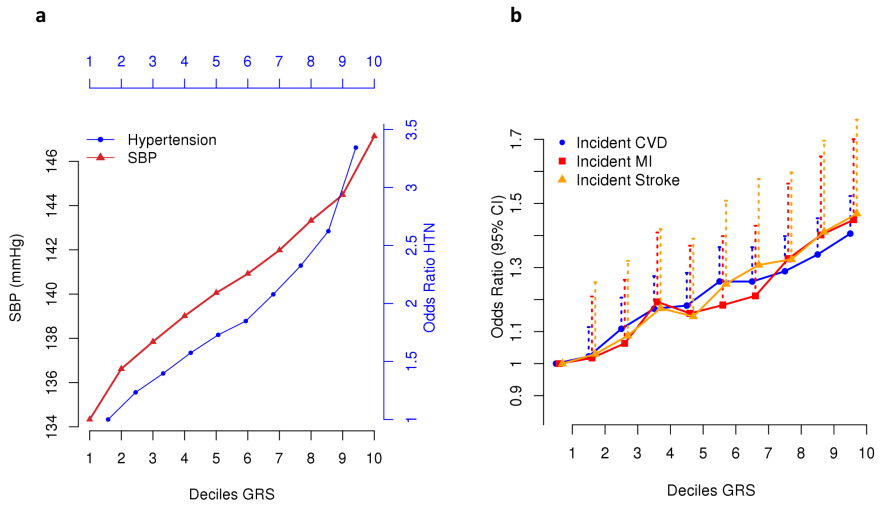
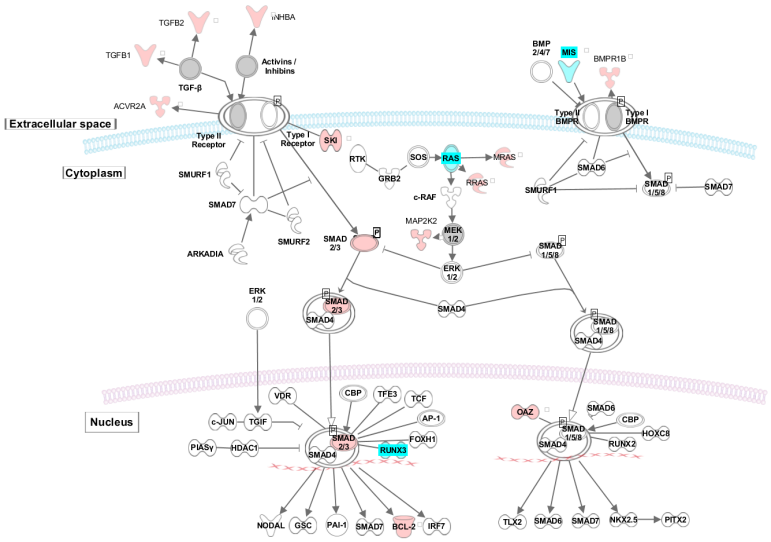


Figure 8: Known and novel BP associations in the TGFβ signalling pathway. Genes with known associations with BP are indicated in cyan. Genes with novel associations with BP reported in this study are indicated in red. TGFβ pathway was derived from an ingenuity canonical pathway. BP: Blood Pressure.



ONLINE METHODS

UK Biobank (UKB) data

We performed a Genome Wide Association Study (GWAS) analysis in 458,577 UKB participants[13] (**Supplementary Methods**). These consist of 408,951 individuals from UKB 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 Axiom Array chip from the UK BiLEVE study[57], which is a subset of UKB. SNPs were imputed centrally by UKB using a reference panel that merged the UK10K and 1000 Genomes Phase 3 panel as well as the Haplotype Reference Consortium (HRC) panel[58]. For current analysis only SNPs imputed from the HRC panel were considered.

UKB phenotypic data

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)**.

Three BP traits were analysed: systolic (SBP), diastolic (DBP) and pulse pressure (PP) (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 single value. After calculating BP values, we adjusted for medication use by adding 15 and 10 mmHg to SBP and DBP, respectively, for individuals reported to be taking BP-lowering medication (94,289 individuals)[59]. Descriptive summary statistics are shown in **Supplementary Table 1a.**

UKB analysis models

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) software[17]. We also calculated the estimated SNP-wide heritability (h^2) in our data. We used genotyped SNPs filtered for MAF > 5%; HWE $P > 1 \times 10^{-6}$; missingness < 0.015, to estimate the parameters of the linear mixed model, for the initial modelling step only. Within the association analysis, we adjust for the following covariates: sex, age, age^2, BMI and a binary indicator variable for UKB vs UK BiLEVE to account for the different genotyping chips. The association analysis performed by BOLT-LMM (v2.3) corrects for population structure and cryptic relatedness in very large datasets[17]. The genome-wide association analysis of all HRC-imputed SNPs was restricted to variants with MAF ≥ 1% and INFO > 0.1.

Genomic inflation and confounding

We applied the univariate LD score regression method (LDSR)[18] to test for genomic inflation (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 standard errors) were 1.217 (0.018), 1.219 (0.020) and 1.185 (0.017) for SBP, DBP and PP respectively, and were used to adjust the UKB GWAS results for genomic inflation, prior to the meta-analysis**.**

International Consortium for Blood Pressure (ICBP) GWAS

ICBP GWAS is an international consortium to investigate BP genetics[6]. 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: ASCOT-SC, ASCOT-UK, BRIGHT, Dijon 3C, EPIC-CVD, GAPP, HCS, GS:SFHS, Lifelines, JUPITER, PREVEND, TWINSUK, GWAS-Fenland, InterAct-GWAS, OMICS-EPIC, OMICS-Fenland, UKHLS, GoDARTS-Illumina and GoDarts-Affymetrix, NEO, MDC, SardiNIA, METSIM.

All study participants were of European ancestry and were imputed to either the 1000 Genomes Project Phase 1 integrated release v.3 [March 2012] all ancestry reference panel[62] or the Haplotype Reference Consortium (HRC) panel[16]. The final enlarged ICBP GWAS dataset included 77 cohorts (N=299,024 individuals).

Definition of phenotype data and GWAS analyses of SBP, DBP and PP were as per previous ICBP protocol for 54 studies[11], extended to the additional 23 studies for which new data were available. Full study names, cohort information and general study methods are included in **Supplementary Table 1b** and in **Supplementary Tables 20a-c**. Genomic control was applied at study-level. The LDSR intercepts (standard error) for the ICBP GWAS meta-analysis were 1.089 (0.012), 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 obtain summary results from the combined UKB and ICBP GWAS, for up to N=757,601 participants and ~7.1 M SNPs with MAF ≥ 1% for variants present in both the HRC-imputed 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 BP variants (see below) in the discovery meta-analysis of UKB and ICBP was 1.090 (0.018), 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. 12).** No further correction was applied to the discovery meta-analysis of UKB and ICBP GWAS.

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 and validated as the sentinel SNP in primary analyses from previous BP genetic association studies. These 357 published SNPs correspond to 274 distinct loci, according to locus definition of: (i) SNPs within ±500kb distance of each other; (ii) SNPs in Linkage Disequilibrium (LD), using a threshold of $r^2 \geq 0.1$, calculated with PLINK (v2.0). We then 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 ($r^2 \geq 0.1$).

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

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.

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:

(i) 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;

(ii) the sentinel SNP is genome-wide significant ($P < 5 \times 10^{-8}$) in the combined meta-analysis of discovery and replication (MVP and EGCUT) (replication, described below);

(iii) the sentinel SNP shows support ($P < 0.01$) in the replication meta-analysis of MVP and EGCUT alone **(Supplementary Methods)**;

(iv) the sentinel SNP has concordant direction of effect between the discovery and the replication meta-analyses;

(v) the sentinel SNP must not be located within any of the 274 previously reported loci described above.

The primary replicated trait was then defined as the BP trait with the most significant association from the combined meta-analysis of discovery and replication (in the case where a SNP was replicated for more than one BP trait.)

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

We considered for follow-up SNPs in loci non-overlapping with previously reported loci according to both an LD threshold at r^2 of 0.1 and a 1Mb interval region, as calculated by PLINK[64]. 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. 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 Methods),** and identify the sentinel SNP (most significant) per locus.

Two-stage design: Replication analysis

We used two independent external data sets for replication **(Supplementary Methods)**. We considered SNPs with MAF ≥ 1% for an independent replication in MVP (max N = 220,520)[14] and in EGCUT Biobank (N=28,742)[15]. This provides a total of N = 249,262 independent samples of European descent available for replication. Additional information on the analyses of the two replication datasets is provided in **Supplementary Methods** and in **Supplementary Table 1c**.

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 dataset.

Two-stage design: Combined meta-analysis of discovery and replication meta-analyses

The meta-analyses were performed within METAL software[63] using fixed effects inverse variance weighted meta-analysis **(Supplementary Methods)**. 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.

One-stage design: Overview

Variants that were looked-up but did not replicate according to the two-stage criteria were 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:

i) the sentinel SNP has $P < 5 \times 10^{-9}$ in the discovery (UKB+ICBP) meta-analysis;
ii) the sentinel SNP shows support ($P < 0.01$) in the UKB GWAS alone;
iii) the sentinel SNP shows support ($P < 0.01$) in the ICBP GWAS alone;
iv) the sentinel SNP has concordant direction of effect between UKB and ICBP datasets;
v) The sentinel SNP must not be located within any of the 274 previously reported loci described above or the recently reported non-replicated loci from Hoffman et al.[9] **(Supplementary Table 21)**.

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 false positive findings. The threshold of $P < 5 \times 10^{-9}$ has been proposed as a more conservative statistical significance threshold, e.g. for whole-genome sequencing-based studies[21] and it was more stringent than the suggested p-value (1×10^{-8}) we obtained from our own calculations for the number of independent SNPs tested in the data.

Selection of variants from the meta-analysis of UKB and ICBP was performed as described above for the two-stage design.

Conditional Analysis

We also 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 loci. We used two different methodological approaches, each using the Genome-wide Complex Traits Analysis (GCTA) software[22]: (i) full "genome-wide conditional analysis" with joint multivariate analysis and stepwise model selection across all three BP traits; and (ii) "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 ≥ 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

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 $r^2 \geq 0.8$ with the sentinel SNPs). At the variant level, we use Variant Effect Predictor (VEP) to 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 evaluate all SNPs in LD ($r^2 \geq 0.8$) with our novel sentinel SNPs for evidence of mediation of expression quantitative trait loci (eQTL) in all 44 tissues using the Genotype-Tissue Expression (GTEx) database, to highlight specific tissue types which show eQTLs for a larger than expected proportion of novel loci. We further seek to identify novel loci with the strongest evidence of eQTL associations in arterial tissue, in particular. A locus is annotated with a given eGene only if the most significant eQTL SNP for the given eGene is in high LD ($r^2 \geq 0.8$) with the sentinel SNP, suggesting that the eQTL signal co-localises with the sentinel SNP.

We annotated nearest genes, eGenes (genes whose expression is affected by eQTLs) and Hi-C interactors with HUVEC, HVMSC and HAEC expression from the Fantom5 project. Genes that had higher then median expression levels in the given cell types were indicated as expressed.

To identify SNPs in the novel loci that have a non-coding functional effect (influence binding 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 combinations[65]. A change of >0.1 in the binding score predicted by DeepSEA for the reference and alternative alleles respectively has been shown to have high true positive rate ~80-95% and low false positive rate ~5-10%, therefore we used this cut-off to find alleles with non-coding functional effect.

We identify potential target genes of regulatory SNPs using long-range chromatin interaction (Hi-C) data from HUVECs[23], aorta, adrenal glands, neural progenitor and mesenchymal stem cell, which are tissues and cell types that are considered relevant for regulating BP[24]. Hi-C data are corrected for genomic biases and distance using the Hi-C Pro and Fit-Hi-C pipelines according to Schmitt et al. (40kb resolution – correction applied to interactions with 50kb-5Mb span)[24]. We find the most significant promoter interactions for all potential regulatory SNPs (RegulomeDB score ≤ 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 annotate the loci.

We then perform overall enrichment testing across all loci. Firstly, we use DEPICT[66] (Data-driven Expression Prioritized Integration for Complex Traits) to identify tissues and cells which are highly expressed at genes within the BP loci. DEPICT uses a large number of microarrays (~78K) to identify cells and tissues where the genes are highly expressed and uses pre-computed GWAS phenotypes to adjust for cofounding sources. Secondly, we use DEPICT to test for enrichment in gene sets associated with biological annotations (manually curated and molecular pathways, phenotype data from mouse KO studies). Using the co-expression data DEPICT calculates a probability for each gene to belong to a given gene set

and uses this to weight the enrichment of the genes present in the tested loci. DEPICT provides a *P*-value of enrichment and false discovery rates adjusted *P*-values for each tissue/cells or gene set tested. 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 filtered by highest significance, $P < 1 \times 10^{-12}$).

Furthermore, to investigate cell type specific enrichment within DNase I sites, we used FORGE, which tests for enrichment of SNPs within DNase I sites in 123 cell types from the Epigenomics Roadmap Project and ENCODE[25] **(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

At the gene level, we use Ingenuity Pathway Analysis (IPA) software (IPA®, QIAGEN 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 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 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

As a secondary analysis, we look up all known and novel BP-associated SNPs in African (7,782) and South Asian (10,322) ancestry samples of UKB. An equivalent GWAS-LMM analysis was performed using BOLT-LMM for each BP trait within each ancestry **(Supplementary Methods).**

Effects on other traits and diseases

We query SNPs against GWAS catalog[26] and PhenoScanner[27], 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 ($r^2 \geq 0.8$) with the 535 sentinel novel SNPs, to highlight the loci with strongest evidence of association with other traits. We further evaluated these effects using DisGeNET, a resource that integrates data from expert curated repositories, GWAS catalogues, animal models and the literature[28,29]. Specifically, at the SNP level, overlaps with DisGeNET terms were computed, with roughly the same number of markers in the published and novel BP loci. Thus, given the expected saturation of the overlaps, a more than double increase indicates that strong associations are more frequent for the novel BP loci. At the gene level, overrepresentation enrichment analysis (ORA) with WebGestalt[67] on the nearest genes to all BP loci was carried out. Moreover, we tested sentinel SNPs at all published and novel (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 ATLAS[68]. 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

We calculated a genetic risk score (GRS) 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. We first create two trait-specific weighted GRSs (i.e. SBP, DBP), for all pairwise-independent, LD-filtered ($r^2 < 0.1$) previously reported variants and 535 novel sentinel variants combined**.** For the previously reported variants, we weight BP increasing alleles by the trait-specific beta coefficients from the new ICBP meta-analysis GWAS that is part of the discovery stage (to minimize winner's curse bias compared to using UKB where the majority of discovery of published SNPs has been derived). For the novel variants, beta coefficients of the replication meta-analysis for each BP trait are used as independent, unbiased weights. We then derive a single BP GRS as the average of the GRS for SBP and DBP, and standardize it to have mean zero and standard deviation of one. GRS were calculated for 487,409 individuals. For statistical analysis of GRS, we focused on unrelated individuals only. The UKB database included 502,638 individuals. We removed pregnant women (n=372) and withdrawn individuals (n=19). We merged GRS and phenotype data (n_{merged}=487,048) and excluded individuals with first or second degree related individuals. The final database for analysis included 423,713 unrelated individuals of European ancestry of whom 392,092 individuals were free of cardiovascular events at baseline.

We assess the association of the continuous GRS variable on BP by simple linear regression, and use logistic regression to examine the association of the GRS with risk of hypertension, with and without adjustment for sex. We then use linear and logistic regression to compare BP levels and risk of hypertension, respectively, for individuals in the top vs bottom quintiles of the GRS 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, we also carried out similar analyses in Airwave, an independent cohort of N=14,004 unrelated participants of European descent[30] **(Supplementary Methods).**

We also assessed the association of the GRS with cardiovascular disease in unrelated participants in UKB data, based on self-reported medical history, and linkage to hospitalization and mortality data. We use logistic regression with binary outcome variables for composite incident cardiovascular disease **(Supplementary Methods)**, incident myocardial infarction and incident stroke (using the algorithmic UKB definitions) and GRS as explanatory variable (with and without sex adjustment).

We also calculated the association of this GRS with BP in unrelated individuals of African (N=6,970) and South Asian (N=8,827) ancestry from the UKB using the approach described above, to see whether BP-associated SNPs identified from GWAS predominantly in Europeans are also associated with BP in populations of non-European ancestry.

We calculated the percentage of variance in BP explained by genetic variants using the independent Airwave cohort (N=14,004). We generated the residuals from a regression of each trait against age, age^2, sex and BMI. We then fit a second linear model for the trait residuals with the GRS plus the top 10 principal components, and estimated the percentage variance of the dependent (BP) variable explained by the GRS. We considered three different levels of the GRS: (i) all pairwise-independent, LD-filtered ($r^2 < 0.1$) published SNPs within the known loci; (ii) all known SNPs and sentinel SNPs at novel loci; (iii) all independent signals at all 901 known and novel loci including the 163 secondary SNPs.

Data availability statement

The genetic and phenotypic UKB data are available upon application to the UK Biobank (https://www.ukbiobank.ac.uk). All replication data 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.

Ethics Statement

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

References

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).
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; Marchini, J. Genome-wide geentic data on 500,000 UK Biobank Participants. *bioRxiv* **166298** (2017).
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).
60. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190 (2017).
61. 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).
62. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

63. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
64. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
65. Zhou, J. & Troyanskaya, O.G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* **12**, 931-4 (2015).
66. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
67. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* (2017).
68. Canela-Xandri, O.R., Konrad; Tenesa, Albert. An atlas of genetic associations in UK Biobank. *bioRxiv* **176834** (2017).