Genome-wide association analyses of >200,000 individuals identify 58 genetic loci for chronic inflammation and highlights pathways that link inflammation and complex disorders

Symen Ligthart[1], Ahmad Vaez[2], Urmo Võsa[3], Maria G Stathopoulou[4], Paul S de Vries[5], Bram P Prins[6], Peter J Van der Most[2], Toshiko Tanaka[7], Elnaz Naderi[2, 8], Lynda M Rose[9], Ying Wu[10], Robert Karlsson[11], Maja Barbalic[12], Honghuang Lin[13], René Pool[14, 15], Gu Zhu[16], Aurélien Macé[17-19], Carlo Sidore[20], Stella Trompet[21, 22], Massimo Mangino[23, 24], Maria Sabater-Lleal[25], John P Kemp[26, 27], Ali Abbasi[2, 28, 29], Tim Kacprowski[30, 31], Niek Verweij[32], Albert V. Smith[33, 34], Tao Huang[35, 36], Carola Marzi[37, 38], Mary F. Feitosa[39], Kurt K Lohman[40], Marcus E Kleber[41], Yuri Milaneschi[42], Christian Mueller[43-45], Mahmudul Huq[2], Efthymia Vlachopoulou[46], Leo-Pekka Lyytikäinen[47, 48], Christopher Oldmeadow[49, 50], Joris Deelen[51, 52], Markus Perola[53], Jing Hua Zhao[54], Bjarke Feenstra[55], LifeLines Cohort Study[2], Marzyeh Amini[2], CHARGE Inflammation Working Group, Jari Lahti[56-58], Katharina E Schraut[59], Myriam Fornage[60], Bhoom Suktitipat[61], Wei-Min Chen[62], Xiaohui Li[63], Teresa Nutile[64], Giovanni Malerba[65], Jian'an Luan[54], Tom Bak[66], Nicholas Schork[67, 68], Fabiola Del Greco M[69], Elisabeth Thiering[70, 71], Anubha Mahajan[72], Riccardo E Marioni[73, 74], Evelin Mihailov[75], Joel Eriksson[76], Ayse Bilge Ozel[77], Weihua Zhang[78, 79], Maria Nethander[80], Yu-Ching Cheng[81], Stella Aslibekyan[82], Wei Ang[83], Ilaria Gandin[84], Loïc Yengo[85], Laura Portas[86], Charles Kooperberg[87], Edith Hofer[88, 89], Kumar B Rajan[90], Claudia Schurmann[91, 92], Wouter den Hollander[93], Tarunveer S Ahluwalia[94, 95], Jing Zhao[96], Harmen HM Draisma[14, 15, 97], Ian Ford[98], Nicholas Timpson[99], Alexander Teumer[100], Hongyan Huang[101], Simone Wahl[37, 38], YongMei Liu[40], Jie Huang[102], Hae-Won Uh[103], Frank Geller[55], Peter K Joshi[59], Lisa R. Yanek[104], Elisabetta Trabetti[65], Benjamin Lehne[78], Diego Vozzi[105], Marie Verbanck[85], Ginevra Biino[106], Yasaman Saba[107], Ingrid Meulenbelt[93], Jeff R O'Connell[81], Markku Laakso[108], Franco Giulianini[9], Patrik KE Magnusson[11], Christie M Ballantyne[109, 110], Jouke Jan Hottenga[14], Grant W Montgomery[16], Fernando Rivadineira[111], Rico Rueedi[17, 18], Maristella Steri[20], Karl-Heinz Herzig[112-114], David J Stott[115], Cristina Menni[23], Mattias Frånberg[25, 116], Beate St Pourcain[99, 117], Stephan B Felix[31, 118], Tune H Pers[55, 95], Stephan JL Bakker[32], Peter Kraft[101], Annette Peters[119], Dhananjay Vaidya[104], Graciela Delgado[41], Johannes H Smit[42], Vera Großmann[120], Juha

Sinisalo[121], Ilkka Seppälä[47, 48], Stephen R Williams[122], Elizabeth G Holliday[49, 50], Matthijs Moed[51], Claudia Langenberg[54],Katri Räikkönen[57], Jingzhong Ding[123], Harry Campbell[59], Michele M Sale[62], Yii-Der I Chen[63], Alan L James[124, 125], Daniela Ruggiero[64, 126], Nicole Soranzo[127], Catharina A Hartman[66], Erin N Smith[128], Gerald S Berenson[129], Christian Fuchsberger[69], Dena Hernandez[130], Carla MT Tiesler[70, 71], Vilmantas Giedraitis[131], David Liewald[73], Krista Fischer[75], Dan Mellström[76], Anders Larsson[132], Yunmei Wang[133], William R Scott[78], Matthias Lorentzon[76, 134], John Beilby[135, 136], Kathleen A Ryan[81], Craig E Pennell[83], Dragana Vuckovic[137], Beverly Balkau[138], Maria Pina Concas[105], Reinhold Schmidt[88], Carlos F Mendes de Leon[139], Erwin P Bottinger[91, 140], Margreet Kloppenburg[141, 142], Lavinia Paternoster[27], Michael Boehnke[143], AW Musk[124, 144], Gonneke Willemsen[14], David M Evans[26, 99], Pamela AF Madden[145], Mika Kähönen[146, 147], Zoltán Kutalik[18, 19], Magdalena Zoledziewska[20], Ville Karhunen[148], Stephen B Kritchevsky[149] Naveed Sattar[150], Genevieve Lachance[23], Robert Clarke[151], Tamara B Harris[152], Olli T Raitakari[153, 154], John R Attia[49, 50, 155], Diana van Heemst[22], Eero Kajantie[156-158], Rossella Sorice[64], Giovanni Gambaro[159], Robert A Scott[54], Andrew A Hicks[69], Luigi Ferrucci[7], Marie Standl[70], Cecilia M Lindgren[72, 160, 161], John M Starr[73, 162], Magnus Karlsson[163], Lars Lind[132], Jun Z Li[77], John C Chambers[78, 79, 164-166], Trevor A Mori[83], Eco JCN de Geus[14, 15], Andrew C Heath[145], Nicholas G Martin[16], Juha Auvinen[148, 167], Brendan B Buckley[168], Anton JM de Craen[22], Melanie Waldenberger[37, 169], Konstantin Strauch[170, 171], Thomas Meitinger[172-174], Rodney J Scott[49, 175], Mark McEvoy[49, 50], Marian Beekman[51], Cristina Bombieri[65], Paul M Ridker[9, 176], Karen L Mohlke[10], Nancy L Pedersen[11], Alanna C Morrison[5], Dorret I Boomsma[14], John B Whitfield[16], David P Strachan[177], Albert Hofman[1], Peter Vollenweider[178], Francesco Cucca[20], Marjo-Riitta Javelin[113, 148, 167, 179], J Wouter Jukema[21, 180, 181], Tim D Spector[23], Anders Hamsten[25], Tanja Zeller[43, 45], André G Uitterlinden[1, 111], Matthias Nauck[31, 182], Vilmundur Gudnason[33, 34], Lu Qi[36, 183], Harald Grallert[37, 38], Ingrid B Borecki[184], Jerome I Rotter[185], Winfried März[41, 186, 187], Philipp S Wild[45, 120, 188], Marja-Liisa Lokki[46], Michael Boyle[155], Veikko Salomaa[53], Mads Melbye[55, 189, 190], Johan G Eriksson[53, 58, 191], James F Wilson[59, 192], Brenda WJH Penninx[42], Diane M Becker[104], Bradford B Worrall[193], Greg Gibson[96], Ronald M Krauss[194], Marina Ciullo[64, 126], Gianluigi Zaza[195], Nicholas J Wareham[54], Albertine J Oldehinkel[66], Lyle J Palmer[196], Sarah S Murray[197], Peter P Pramstaller[69, 198, 199], Stefania Bandinelli[200],

Joachim Heinrich[70], Erik Ingelsson[201, 202], Ian J Deary[73, 203], Reedik Magi[75], Liesbeth Vandenput[76], Pim van der Harst[32], Karl C Desch[204], Jaspal S Kooner[79, 164, 166, 205], Claes Ohlsson[76], Caroline Hayward[192], Terho Lehtimäki[47, 48], Alan R Shuldiner[81], Donna K Arnett[206], Lawrence J Beilin[83], Antonietta Robino[105], Philippe Froguel[85, 207], Mario Pirastu[86], Tine Jess[208], Wolfgang Koenig[169, 209, 210], Ruth JF Loos[91, 92, 211], Denis A Evans[90], Helena Schmidt[107, 212], George Davey Smith[99], P Eline Slagboom[51], Gudny Eiriksdottir[33], Andrew P Morris[72, 213], Bruce M Psaty[214-217], Russell P Tracy[218], Ilja M Nolte[2], Eric Boerwinkle[219, 220], Sophie Visvikis-Siest[4], Alex P Reiner[215], Myron Gross[221], Joshua C Bis[214], Lude Franke[3], Oscar H Franco[1], Emilia J Benjamin[13, 222], Daniel I Chasman[9, 176], Josée Dupuis[222, 223], Harold Snieder[2], Abbas Dehghan[179]*, Behrooz Z Alizadeh[2]*

Anton JM de Crean deceased.

Affiliations

1. Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3000 CA, the Netherlands.
2. Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen 9700 RB , the Netherlands.
3. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen 9700 RB, the Netherlands.
4. UMR INSERM U1122; IGE-PCV "Interactions Gène-Environnement en Physiopathologie Cardio-Vasculaire", Université de Lorraine, Nancy 54000, France.
5. Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.
6. Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK.
7. Translational Gerontology Branch, National Institute on Aging, Baltimore, MD 21224, USA.
8. Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, Groningen 9713 GZ, the Netherlands.
9. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA.
10. Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA.
11. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden.
12. University of Split School of Medicine, Split 21000, Croatia.
13. Department of Medicine, Boston University School of Medicine, Boston, MA 01702, USA.
14. Department of Biological Psychology, Netherlands Twin Register, Vrije Universiteit, Amsterdam 1081 BT, the Netherlands.
15. Amsterdam Public Health research institute, VU University medical center, Amsterdam 1081 BT, the Netherlands.
16. QIMR Berghofer Medical Research Institute, Brisbane QLD 4006, Australia.
17. Department of Computational Biology, University of Lausanne, Lausanne 1010, Switzerland.
18. Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland.
19. Institute of Social and Preventive Medicine, University Hospital of Lausanne, Lausanne 1010, Switzerland.
20. Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Sardinia 08045, Italy.
21. Department of Cardiology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands.
22. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2333 ZA, the Netherlands.

23. Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK.
24. NIHR Biomedical Research Centre at Guy's and St. Thomas' Foundation Trust, London SE1 9RT, UK.
25. Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Stockholm 17176, Sweden.
26. University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland QLD 4102, Australia.
27. MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK.
28. Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, the Netherlands.
29. MRC Epidemiology Unit, University of Cambridge School of Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.
30. Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt University Greifswald, Greifswald 17475, Germany.
31. DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald 17475, Germany.
32. Department of Cardiology and Thorax Surgery - Experimental Cardiology, University of Groningen, University Medical Center Groningen, Groningen 9713 AV, the Netherlands.
33. Icelandic Heart Association, Kopavogur, 201, Iceland.
34. Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland.
35. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China.
36. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA.
37. Institute of Epidemiology II, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany.
38. German Center for Diabetes Research (DZD e.V.), Partner Munich, Munich 85764, Germany.
39. Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108-2212, USA.
40. Department of Epidemiology and Prevention, Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27157, USA.
41. Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim 68167, Germany.
42. Department of Psychiatry, EMGO Institute for Health and Care Research and Neuroscience Campus Amsterdam, VU University Medical Center/GGZ inGeest, Amsterdam 1081 BT, The Netherlands.
43. Department of General and Interventional Cardiology, University Heart Center Hamburg, Hamburg 20246, Germany.
44. Institute of Medical Biometry and Statistics, University Medical Center Schleswig-Holstein, Campus Luebeck, Lübeck 23562, Germany.
45. German Center for Cardiovascular Research (DZHK e.V.), partner site Hamburg, Lübeck, Kiel 20246, Germany.
46. Transplantation Laboratory, Medicum, University of Helsinki, Helsinki 00014, Finland.
47. Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33014, Finland.
48. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland.
49. Hunter Medical Research Institute, New Lambon Heights, NSW 2305, Australia.
50. Centre for Clinical Epidemiology & Biostatistics (CCEB), Faculty of Health and Medicine, University of Newcastle, Callaghan, NSW 2308, Australia.
51. Molecular Epidemiology, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands.
52. Max Planck Institute for Biology of Ageing, Cologne 50931, Germany.
53. National Institute for Health and Welfare, Helsinki 00271, Finland.
54. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge CB2 0QQ, UK.
55. Department of Epidemiology Research, Statens Serum Institut, Copenhagen 2300, Denmark.
56. Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki 00014, Finland.
57. Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki 00014, Finland.
58. Folkhälsan Research Centre, Helsinki 00250, Finland.

59. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH16 4UX, Scotland.
60. Human Genetics Center, School of Public Health and Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA.
61. Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.
62. Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA.
63. Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.
64. Institute of Genetics and Biophysics "A. Buzzati-Traverso" - CNR, Napoli 80131, Italy.
65. Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona 37134, Italy.
66. Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), University Medical Center Groningen, University of Groningen, Groningen 9700 RB, the Netherlands.
67. Human Biology, J. Craig Venter Institute, La Jolla, CA 92037, USA.
68. Quantitative Medicine, The Translational Genomics Research Institute (Tgen), Phoenix, AZ 85004, USA.
69. Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, 39100, Italy.
70. Institute of Epidemiology I, Helmholtz Zentrum München – German Research Centre for Environmental Health, Neuherberg 85764, Germany.
71. Ludwig-Maximilians-University of Munich, Dr. von Hauner Children's Hospital, Munich 80337, Germany.
72. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.
73. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK.
74. Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
75. Estonian Genome Center, University of Tartu, Tartu 51010, Estonia.
76. Department of Internal Medicine and Clinical Nutrition, University of Gothenburg, Gothenburg 41345, Sweden.
77. Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109-5618, USA.
78. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK.
79. Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK.
80. Bioinformatics Core Facility, The Sahlgrenska Academy, University of Gothenburg, Gothenburg 413 90, Sweden.
81. Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore 21201, MD, USA.
82. Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294-0022, USA.
83. Medical School, The University of Western Australia, Perth, WA 6009, Australia.
84. AREA Science Park, Trieste 34149, Italy.
85. CNRS UMR 8199 - University of Lille - Institut Pasteur de Lille, European Genomic Institute for Diabetes (EGID), FR 3508, F-59000 Lille, France.
86. Support OU, Institute of Genetic and Biomedic Research, National Research Council of Italy (CNR), Sassari 7100, Italy.
87. Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Mail Stop M3-A410, 1100 Fairview Ave. N., Seattle, WA, USA.
88. Clinical Division of Neurogeriatrics, Department of Neurology, Medical University Graz, Graz 8036, Austria.
89. Institute of Medical Informatics, Statistics and Documentation, Medical University Graz, Graz 8036, Austria.
90. Department of Internal Medicine, Rush University Medical Center, Chicago, IL 60612, USA.
91. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
92. The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
93. Department of Medical Statistics and Bio-informatics, Section Molecular Epidemiology, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands.

94. Steno Diabetes Center Copenhagen, Gentofte 2820, Denmark.
95. Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2100, Denmark.
96. Center for Integrative Genomics, School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA.
97. Neuroscience Campus Amsterdam, Amsterdam 1081 HV, the Netherlands.
98. Robertson Center for Biostatistics, University of Glasgow, Glasgow G12 8QQ, UK.
99. MRC Integrative Epidemiology Unit, University of Bristol, UK.
100. Department SHIP-KEF, Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany.
101. Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA 2115, USA.
102. Boston VA Research Institute, Inc., Boston, MA 02130, USA.
103. Medical statistics and bioinformatics, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands.
104. GeneSTAR Research Center, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA.
105. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste 34140, Italy.
106. Institute of Molecular Genetics, National Research Council of Italy (CNR), Pavia 27100, Italy.
107. Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz 8010, Austria.
108. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio 70210, Finland.
109. Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA.
110. Methodist DeBakey Heart and Vascular Center, Houston, TX 77030, USA.
111. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam 3015 CN, the Netherlands.
112. Department of Physiology, Institute of Biomedicine, University of Oulu, Medical Research Center Oulu and Oulu University Hospital, Oulu 90014, Finland.
113. Biocenter Oulu, University of Oulu, Oulu 90220, Finland.
114. Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan 60-512, Poland.
115. Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow G12 8QQ, UK.
116. Department of Numerical Analysis and Computer Science, Stockholm University, Stockholm 100 44, Sweden.
117. Donders Institute, Radboud University, Nijmegen 6525 XD, the Netherlands.
118. Department for Internal Medicine B, University Medicine Greifswald, Greifswald 17475, Germany.
119. Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherber 85764, Germany.
120. Center for Thrombosis and Hemostasis (CTH), University Medical Center of the Johannes Gutenberg University Mainz, Mainz 55131, Germany.
121. Heart and Lung Center, Helsinki University Hospital and Helsinki University, Helsinki 00029, Finland.
122. Department of Neurology, University of Virginia, Charlottesville, VA 22908, USA.
123. Department of Internal Medicine/Geriatrics, Wake Forest University Health Sciences, Winston-Salem, NC 27157, USA.
124. Busselton Population Medical Research Institute, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
125. Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
126. IRCCS Neuromed, Pozzilli (IS) 86077, Italy.
127. Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK.
128. Department of Pediatrics and Rady Children's Hospital, University of California at San Diego, School of Medicine, La Jolla, CA 92037, USA.
129. Center for Cardiovascular Health, Tulane University, New Orleans, LA 70112, USA.
130. Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA.

131. Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala 752 37, Sweden.
132. Department of Medical Sciences, Uppsala University, Uppsala 751 41, Sweden.
133. Department of Medicine, Case Cardiovascular Research Institute, Case Western Reserve University; Harrington Heart & Vascular Institute, University Hospitals, Cleveland, OH 44106, USA.
134. Geriatric Medicine, Sahlgrenska University Hospital, Mölndal 431 80, Sweden.
135. PathWest Laboratory Medicine WA, Nedlands, WA 6009, Australia.
136. School of Pathology and Laboratory, The University of Western Australia, Crawley, Perth, WA 6009, Australia.
137. Medical Sciences, Surgical and Health Department, University of Trieste, Trieste 34137, Italy.
138. Inserm U-1018, CESP, Team 5 (EpReC, Renal and cardiovascular Epidemiology), UVSQ-UPS, Villejuif, 94807, France.
139. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.
140. Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
141. Department of Rheumatology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands.
142. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands.
143. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA.
144. Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
145. Department of Psychiatry, Washington University School of Medicine, 4560 Clayton Ave., Suite 1000, St. Louis, MO, 63110, USA.
146. Department of Clinical Physiology, Tampere University Hospital, Tampere 33520, Finland.
147. Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland.
148. Center for Life Course Epidemiology, Faculty of Medicine, University of Oulu, Oulu 90014, Finland.
149. Gerontology and Geriatric Medicine, Sticht Center on Aging and Rehabilitation, Wake Forest University Health Sciences, Winston-Salem, NC 27157, USA.
150. BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow G12 8TA, UK.
151. Clinical Trial Service Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK.
152. Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, Maryland 20892, USA.
153. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland.
154. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland.
155. John Hunter Hospital, New Lambton Heights, NWS 2305, Australia.
156. Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki 00014, Finland.
157. Hospital for Children and Adolescents, Helsinki University Central Hospital and University of Helsinki, Helsinki 00290, Finland.
158. Department of Obstetrics and Gynaecology, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu 90014, Finland.
159. Division of Nephrology and Dialysis, Columbus-Gemelli University Hospital, Università Cattolica del Sacro Cuore, Roma 168, Italy.
160. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7FZ, UK.
161. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA.
162. Alzheimer's Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK.
163. Department of Clinical Sciences and Orthopaedic Surgery, Lund University, Malmo 20502, Sweden.
164. Imperial College Healthcare NHS Trust, London W12 0HS, UK.
165. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore.
166. MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK.
167. Unit of Primary Care, Oulu University Hospital, Oulu 90220 , Finland.
168. Department of Pharmacology and Therapeutics, University College Cork, Cork T12 YT20, Ireland.

169. DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich 80336, Germany.
170. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg 85764, Germany.
171. Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Neuherberg 85764, Germany.
172. Institute of Human Genetics, Technische Universität München, Munich 85764, Germany.
173. Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg 85764, Germany.
174. Munich Cluster for Systems Neurology (SyNergy), Munich 81377, Germany.
175. Information-based Medicine Stream, Hunter Medical Research Institute, New Lambton Heights, NSW 2305, Australia.
176. Harvard Medical School, Boston, MA 02115, USA.
177. Population Health Research Institute, St George's, University of London, London SW17 0RE, UK.
178. Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne 1011, Switzerland.
179. Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London W2 1PG, UK
180. Durrer Center for Cardiogenetic Research, Amsterdam 3501 DG, the Netherlands.
181. Interuniversity Cardiology Institute of the Netherlands, Utrecht 3511 EP, the Netherlands.
182. Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald 17475, Germany.
183. Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA 70112, USA.
184. Analytical Genetics Group, Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY 10591, USA.
185. Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.
186. Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim 68161, Germany.
187. Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036, Austria.
188. Preventive Cardiology and Preventive Medicine, Center for Cardiology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz 55131, Germany.
189. Department of Clinical Medicine, University of Copenhagen, Copenhagen 2300, Denmark.
190. Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
191. Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki 00014, Finland.
192. MRC Human Genetics Unit, Institute for Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, Scotland.
193. Departments of Neurology and Public Health Sciences, University of Virginia Charlottesville, Charlottesville, VA 22908-0394, USA.
194. Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA
195. Renal Unit, Department of Medicine, Verona University Hospital, Verona 37126, Italy.
196. School of Public Health, University of Adelaide, Adelaide SA 5000, Australia.
197. Department of Pathology, University of California San Diego, San Diego, CA 92121, USA.
198. General Central Hospital, Department of Neurology, Bolzano 39100, Italy.
199. Department of Neurology, University of Lübeck, Lübeck 23538, Germany.
200. Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence 50122, Italy.
201. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala 751 41, Sweden.
202. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
203. Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK.
204. Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI 48109, USA.
205. National Heart and Lung Institute, Imperial College London, London W12 0NN, UK.
206. University of Kentucky, College of Public Health, Lexington, KY 40508, USA.

207. Department of Genomics of Common Disease, School of Public Health, Imperial College London, London SW7 2AZ, UK.
208. Department of Clinical Epidemiology, Bispebjerg and Frederiksberg Hospital, Frederiksberg 2200, Denmark.
209. Department of Internal Medicine II-Cardiology, University of Ulm Medical Center, Ulm 80636, Germany.
210. Deutsches Herzzentrum München, Technische Universität München, Munich 80636, Germany.
211. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029-6542, USA.
212. Department of Neurology, Medical University Graz, Graz 8010, Austria.
213. Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK.
214. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA.
215. Department of Epidemiology, University of Washington, Seattle, WA 98101, USA.
216. Department of Health Services, University of Washington, Seattle, WA 98195-7660, USA.
217. Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA.
218. Department of Pathology, University of Vermont, Colchester, VT 05405, USA.
219. Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.
220. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.
221. Department of Lab Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA.
222. National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA 01702, USA.
223. Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA.

***Corresponding authors**

Abbas Dehghan MD PhD

Department of Epidemiology & Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, UK

Room 157, Norfolk Place, St Mary´s Campus

Phone: 0044 20 7594 3347

Email: a.dehghan@imperial.ac.uk

Behrooz Z. Alizadeh MD PhD

Department of Epidemiology, University Medical Center Groningen

Hanzeplein 1, 9713 GZ Groningen, the Netherlands

Phone: 0031 50 361 0738, Fax: 0031 50 361 4493

Email: b.z.alizadeh@umcg.nl

Abstract

C-reactive protein (CRP) is a sensitive biomarker of chronic low-grade inflammation that is associated with multiple complex diseases. The genetic determinants of chronic inflammation remain largely unknown, and the causal role of CRP in several clinical outcomes is debated. We performed two genome-wide association studies (GWAS), HapMap and 1000Genomes imputed, of circulating CRP levels using data from 88 studies comprising 204,402 European individuals. Additionally, we performed *in silico* functional analyses and Mendelian randomization analyses with several clinical outcomes. The GWAS meta-analyses of CRP revealed 58 distinct genetic loci ($P<5\times10^{-8}$). After adjustment for body mass index in the regression analysis, the associations at all except three loci remained. The lead variants at the distinct loci explained up to 7.0% of the variance in circulating CRP levels. We identified 66 gene sets that were organized in two substantially correlated clusters, one mainly comprised of immune pathways, and the other characterized by metabolic pathways in the liver. Mendelian randomization analyses revealed a causal protective effect of CRP on schizophrenia and a risk increasing effect on bipolar disorder. Our findings provide further insights in the biology of inflammation that may lead to interventions to treat inflammation and its clinical consequences.

Introduction

Inflammation plays a key role in the development of complex diseases such as cardiovascular disease[1], type 2 diabetes[2], Alzheimer's disease[3], and schizophrenia[4]. C-reactive protein (CRP) is a sensitive marker of chronic low-grade inflammation[5], and elevated serum levels of CRP have been associated with a wide range of diseases[6-8]. Unraveling the genetics of inflammation may provide further insights into the underlying biology of inflammation, and may identify therapeutic targets to attenuate inflammation.

The genetic determinants of CRP have only been partly characterized. In 2011, our group published a HapMap-based meta-analysis of genome-wide association studies (GWAS) including a discovery panel of up to 65,000 individuals and found 18 loci that were associated with CRP levels[9]. Increasing study sample size in GWAS and denser mapping of the genome with further advanced imputation panels may help to identify further genes associated with the phenotypes of interest[10; 11]. Furthermore, by using genetic instrumental variables (i.e., a genetic score), Mendelian randomization (MR) allows investigation of the potential causal effect of an exposure on clinical outcomes, and may help to understand the causal pathways that link the exposure with the outcome[12]. The causal role of CRP in the development of diseases is still controversial[13], and the causal pathways that link inflammation to complex disorders are only partly understood.

We applied two large-scale GWAS on circulatory levels of CRP using HapMap and 1000Genomes (1KG) imputed data to identify genetic determinants of chronic inflammation. Because body mass index (BMI) is a major determinant of CRP levels, we additionally conducted GWAS adjusted for BMI to identify associated loci independent of BMI. To identify any sex differences in genetic determinants of chronic inflammation, we further conducted GWAS in men and women separately. We applied *in silico* functional analyses on the identified loci to obtain better insights into the biological processes potentially regulating chronic inflammation. Finally, we conducted MR analyses to provide an improved understanding of the causal relation between CRP and several related clinical outcomes.

Material and methods

GWAS for circulating CRP levels

We conducted a meta-analysis of GWAS including individuals of European ancestry within the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium Inflammation Working Group of the (CIWG)[14]. The CIWG invited cohorts for participation in the HapMap GWAS meta-analysis of CRP levels in 2012. In 2014 and in light of our assessment which showed complementary value of HapMap and 1KG imputed GWAS[10], we invited studies to participate in the 1KG GWAS meta-analysis. The 1KG GWAS may help to identify loci that were not covered in the HapMap GWAS and fine map loci found in the HapMap GWAS. Cohorts were allowed to participate in either the HapMap or 1KG GWAS, or both. Here we present both a HapMap (204,402 individuals from 78 studies) and 1KG (148,164 individuals from 49 studies) imputed genotypes GWAS meta-analysis. All participating cohorts implemented a pre-specified study plan comprising study design, data quality check, data analysis, and data sharing. We measured serum CRP in mg/L using standard laboratory techniques (Supplemental Methods), and natural log-transformed the values. Individuals with auto-immune diseases, individuals taking immune-modulating agents (if this information was available), and individuals with CRP levels 4SD or more away from the mean were excluded from all analyses. The characteristics of the participants are presented in Table S1. We filtered individuals and genetic variants based on study-specific quality control criteria (Table S2). At each individual study site, we tested genetic variants for association with CRP levels using an additive linear regression model adjusted for age, sex, and population substructure, and accounting for relatedness, if relevant. Before meta-analysis, we filtered variants based on imputation quality at R^2 index of >0.4. To avoid type-I error inflation, we corrected study-specific GWAS for genomic inflation. For the HapMap study, we conducted fixed effect meta-analyses for each genetic variant, using the inverse variance-weighted method implemented in GWAMA[15], and performed a second genomic control on the meta-analysis level. For the 1KG imputed GWAS, we used METAL[16] to perform a fixed effect meta-analysis. We removed variants that were only available in <50% of the samples. The HapMap meta-analysis included 2,254,727 variants, and the 1KG GWAS included 10,019,203 variants. We considered

associations with $P<5\times10^{-8}$ genome-wide significant. We used a stringent distance criterion, 500kb minimum between two significant variants, to identify distinct loci. In each locus, the variant with the smallest p-value was called the "lead variant". Additionally, we performed sex-stratified analyses among HapMap imputed studies, and we tested for heterogeneity between sex-specific effect estimates as described previously[17], using the false-discovery rate of Benjamini-Hochberg to assess significance of the P for sex difference (<0.05). We conducted BMI adjusted analyses in the 1KG meta-analysis to determine the role of BMI in mediating the genetic associations with CRP, and to increase power to detect associations not mediated by BMI.

LD Score regression

Because population stratification is a major concern in GWAS and may lead to false-positive associations, we applied Linkage Disequilibrium Score regression (LDSC) to distinguish whether the inflation of test statistics observed in the CRP GWAS is due to the polygenic architecture of CRP or reflects confounding bias due to cryptic relatedness or population stratification. The LD Score measures collective genetic variation acquired from all genetic variants in LD with the index tagging (causal) variant[18]. A higher LD score of an index variant implicates more nearby genetic variants in high LD with the index variant, which makes it more likely that the index variant tags causal variant(s). More genetic variants in LD with the index genetic variant (i.e., a higher LD score due to polygenicity) may yield higher (i.e., inflated) test statistics. In contrast, higher test statistics caused by cryptic population stratification will not be related to LD score. LD Score regression analysis performs regression of the summary statistics from the GWAS meta-analysis (χ^2 statistics from the GWAS) on the LD scores across the genome. An intercept of the LD Score regression that equals one suggests no confounding bias, whereas an inflated intercept (larger than one) suggests contribution of confounding due to relatedness to the test statistics. We used the LDHub web interface to perform LD Score regression[19]. We filtered variants to the subset of HapMap 3 variants, and excluded variants with duplicated rs numbers, ambiguous variants, minor allele frequency

(MAF)<0.01, and reported sample size <66.7% of total sample size. We used the default European LD Score file based on the European 1KG reference panel.

Furthermore, we applied cross-trait LD score regression to estimate genetic correlation of chronic inflammation (using the HapMap GWAS meta-analysis) with other phenotypes using published GWAS summary statistics[20]. In brief, the cross-product of two GWAS test statistics is calculated at each genetic variant, and this cross-product is regressed on the LD Score. The slope of the regression is used to estimate the genetic covariance between two phenotypes.

Identification of additional distinct variants in associated loci

To identify additional distinct variants in the associated loci, we performed joint approximate conditional analysis using the 1KG meta-analysis summary statistics and the linkage disequilibrium (LD) matrix derived from the first cohort of the Rotterdam study (RS-I) (n=5,974). We used the Genome-wide Complex Trait Analysis (GCTA) tool, which performs a genome-wide step-wise procedure to identify variants according to their distinct association with CRP (i.e., conditional P)[21; 22]. We only used variants with an imputation quality of R^2>0.8 in the reference set (RS-I). This approximate conditional analysis may reveal different lead signals in a locus where multiple associated variants are in the final joint association model. We tested the distinct variants identified in the *CRP* gene jointly for an association with CRP using individual level data from the second and third cohort of the Rotterdam Study (RS-II and RS-III, totaling 5,024 subjects), and the Women's Genome Health Study (WGHS) of 16,299 individuals.

Proportion of CRP variance explained

We estimated the variance explained in serum CRP levels using the formula *(2*MAF(1-MAF)beta^2)/var(CRP)*, where beta is the estimated effect of the individual variants on CRP[23] and var(CRP) is the variance in natural log-transformed CRP estimated in the RS-I cohort. We calculated the variance explained for four combinations of associated variants: 1. the lead variant at just the *CRP* locus; 2. the distinct variants at the CRP locus derived from the 1KG joint conditional analysis; 3. all lead

variants in the distinct loci; 4. all lead variants in the distinct loci and, when applicable, the distinct variants at associated loci derived from the approximate joint conditional analysis.

Pathway analysis and gene expression

We used Data-Driven Expression-Prioritized Integration for Complex Traits[24] (DEPICT v.1 rel173 beta) to systematically prioritize the most likely causal genes, highlight the pathways that are enriched by the likely causal genes and identify tissues and cell types in which genes from associated loci are highly expressed. DEPICT requires summary statistics from the GWAS meta-analysis. First, we filtered genome-wide associated variants from both GWAS meta-analyses by MAF>0.01, and selected variants with low correlation with other variants by PLINK (version 1.90) using a clumping distance of 500 kb apart and/or index of LD r^2 threshold <0.1. The settings for the analysis involved the usage of 1KG pilot phase data[25] (phase 1 integrated release, version 3, CEU, GBR, TSI unrelated individuals; 2010.11.23) with r^2>0.5 LD threshold for locus definition, 10,000 permutations for bias correction, and 500 repetitions for FDR calculation. To summarize and visualize the results, we calculated pairwise Pearson correlation coefficients between all gene-specific Z-scores for every pair of reconstituted DEPICT gene sets. We used Affinity Propagation Clustering (apcluster command; *APCluster* R package[26]) to identify clusters and representative examples of the clusters, and Cytoscape v3.2.1 for visualization of the results. The DEPICT results of the pathway and gene prioritization are summarized as a heatmap (R. v2.3.3, *pheatmap* v1.0.8 package[27]). The gene-specific Z-score describes the likelihood that a given gene is part of the corresponding GO term, KEGG pathway, REACTOME pathway, Mouse Phenotype, or protein-protein interaction network.

Also, we performed Multi-marker Analysis of GenoMic Annotation (MAGMA)[28]. MAGMA performs gene and gene-set analysis and requires the association results of all variants, therefore we chose the larger HapMap GWAS for MAGMA. We used the Functional Mapping and Annotation (FUMA)[29] tool to perform MAGMA, and applied standard settings for running MAGMA.

To prioritize the most likely trait-relevant gene for each GWAS locus, we run colocalization analysis using the "coloc" R package v3.1[30] separately for the HapMap and 1KG GWAS. We used publicly available genome-wide eQTL data from 5,311 whole blood samples[31], and from the Genome Tissue Expression (GTEx) V6p portal incorporating eQTL data from 44 post-mortem tissues[32]. "Coloc" uses approximate Bayes factors to estimate the posterior probability that GWAS and eQTL effects share a single causal variant. All significant cis-eGenes or cis-eProbes ($q<0.05$ in GTEx; lowest cis-eQTL FDR<0.05 in Westra et al.[31]) were extracted ±1Mb from the lead SNP of each locus. The HapMap SNP positions were converted to hg19 with the liftOver command from the rtracklayer v1.38.3 package. We used the SNPs present in both the GWAS and eQTL datasets. For the HapMap GWAS, the 1KG GWAS and the GTEx eQTL datasets, we performed the test using association beta, standard error of beta, and minor allele frequency (MAF). For the data from Westra et al.[31], we used association P-value, MAF, and sample size, and included only the subset of cis-eQTLs which are publicly available (up to significance FDR<0.5). We used default priors supplied by the coloc package (P1=1e-4, P2=1e-4, P12=1e-5; prior probabilities for association in GWAS, eQTL, and both datasets). Full MAF data were not available for the eQTL datasets, therefore we used the GIANT 1KG p1v3 EUR reference panel instead. We visualized the results as a heatmap using the *pheatmap* v1.0.8 R package.[27].

Mendelian randomization

To assess the effect of CRP on complex disorders, we performed a two-sample Mendelian randomization (MR) study on nine clinical outcomes (Alzheimer's disease (AD), bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), inflammatory bowel disease (IBD), rheumatoid arthritis (RA), schizophrenia, and diastolic (DBP) and systolic blood pressure (SBP)) to which CRP showed a potentially causal association at a P<0.1 in a previous MR study[13]. We used the effect estimates of the 48 lead SNPs found to be associated with CRP in the HapMap GWAS, and the effect estimates of the four SNPs that were additionally found to be associated with CRP in the 1KG GWAS in a multiple instrument approach for the MR analyses (n=52 SNPs). Additionally, we separately studied the effect of rs2794520 at the *CRP*

locus to minimize the probability of horizontal pleiotropy that may be introduced in a multiple instrument approach. We tested the statistical significance of the association between the instrument and CRP using the formula:

$$F = \frac{R^2(n-1-k)}{(1-R^2) \times k}$$

R^2 is the variance explained of CRP by the genetic instrument (0.014 for the rs2794520 SNP and 0.065 for the 52-SNP score), n is the number of individuals included in the CRP GWAS, and k the number of variants included in the genetic score. The *F* statistic for the 52-SNP score was 273, and for the rs2794520 SNP 2,902, indicating that both instruments were strong.

For the clinical outcomes, we used summary statistics from the most recent meta-analysis of GWA studies. For diastolic and systolic blood pressure, we used data from the UK Biobank. The details of the outcome studies are summarized in Table S12. We implemented four different methods of MR analyses: Inverse-variance weighted method (IVW), MR-Egger, Weighted median (WM), and Penalized weighted median (PWM). We used the "*TwoSampleMR"* package in R for the MR analyses[33]. Further, we applied the Bonferroni method to correct for multiple testing (0.05/9 phenotypes = 5.6×10^{-3}). When the Q-statistic of the IVW analyses provided evidence for heterogeneity, the weighted median estimates were used for significance. The MR methods are described briefly below.

Inverse-variance weighted (IVW): The causal estimate is obtained by regressing the SNP associations with the outcome on the SNP associations with the risk factor, with the intercept set to zero and weights being the inverse-variances of the SNP associations with the outcome. With a single genetic variant, the estimate is the ratio of coefficients betaY/betaX and the standard error is the first term of the delta method approximation betaYse/betaX. When all CRP-SNPs are valid IVs, the IVW estimates converge to the true causal effect. When one or more invalids IVs are present, (ie. one SNP has effect on outcome through a different pathway than CRP), the IVW estimate deviates from the true causal effect.

MR-Egger: We used MR-Egger to account for potential unbalanced pleiotropy in the multiple variant instrument[34]. When unbalanced pleiotropy is present, an alternative effect (positive or negative) is

present between the SNP and the outcome that may bias the estimate of the causal association. The MR-Egger method is similar to the IVW analysis, but does not force the intercept to pass through the origin. The slope of the MR-Egger regression provides the estimate of the causal association between CRP and the clinical outcome. An MR-Egger intercept that is significantly different from zero suggests directional pleiotropic effects that may bias uncorrected estimates of the causal effect. MR-Egger regression depends on the InSIDE (Instrument Strength Independent of Direct Effect) assumption, that states that the strengths of the effect of the SNP on the outcome is uncorrelated with the direct pleiotropic effect of the SNP on the outcome.

Weighted median (WM) and penalized Weighted Median (PWM): We applied the median based method to provide robust estimates of causal association even in the presence of horizontal pleiotropy when up to 50% of the information contributed by the genetic variants is invalid[35]. In PWM analysis the effect of each variants is weighted by a factor that corresponds to the Q statistics (heterogeneity test) of the SNP; this means that most variants will not be affected by this correction, but the causal effect of the outlying variants, which are most likely to be invalid IVs, will be down-weighted.

We displayed the individual SNP causal effect estimates and corresponding 95% confidence intervals in a forest plot. To assess whether one of the variants used in the genetic score had disproportionate effects, we performed "leave-one-out" analyses where one SNP at a time is removed from the score. We depicted the relationship between the SNP effect on CRP and the SNP effect on the clinical outcomes in a scatter plot, and plotted the individual SNP effect against the inverse of their standard error in a funnel plot. When unbalanced pleiotropy is absent, the causal effect estimates of the individual should center around the meta-analysis estimate in the funnel plot.

We used the proportion of variance in CRP explained by the genetic instruments (0.014 for the rs2794520 SNP and 0.065 for the 52-SNP score) to perform power calculations for each outcome using the online tool mRnd[36]. We calculated the power to detect a relative 5%, 10%, 15%, and 20% difference in outcome risk. For example, a 10% difference refers to an OR of at least 0.90 or 1.10 in outcome risk (Table S13).

Results

HapMap GWAS meta-analysis for CRP levels

The HapMap meta-analysis identified 3,977 genome-wide significant variants at $P<5\times10^{-8}$ (QQ-plot Figure S1; Manhattan plot Figure S2), which mapped to 48 distinct loci (Table 1, Table S3). Of the previously reported 18 variants for CRP, 16 remained associated. Compared to the previous GWAS, the rs6901250 variant at the *GPRC6A* locus (P=0.09) and the rs4705952 variants at the *IRF1* locus ($P=2.7\times10^{-3}$) were not significant. The beta estimates for natural log-transformed CRP for each of the associated loci ranged from 0.020 to 0.229. We observed the strongest association for rs2794520 at the *CRP* gene (β=0.182 in the natural log-transformed CRP (mg/L) per copy increment in the coded allele, $P=4.17\times10^{-523}$), followed by rs4420638 at the *APOC1/E* gene (β=0.229, $P=1.23\times10^{-305}$). Similarly to previous GWAS meta-analysis, the lead variant within the interleukin-6 receptor gene (*IL6R*) was rs4129267 (β=0.088, $P=1.2\times10^{-129}$). Related to the interleukin-6 pathway, we identified rs1880241 upstream of the *IL6* gene (β= 0.028, $P=8.4\times10^{-14}$). In addition to the previously described interleukin-1 signaling, the *IL1RN-IL1F10* locus (interleukin-1 receptor antagonist and interleukin-1 family member 10), we found rs9284725 within the interleukin-1 receptor 1 gene (*IL1R1*, β=0.02, $P=7.3\times10^{-11}$, Table 1**)**. The sex-specific meta-analyses did not identify additional loci for CRP compared to the overall meta-analysis including both sexes, but at four genetic variants we found evidence for heterogeneity in effect estimates between sexes (Table S4), though the directions of associations were consistent.

1KG GWAS meta-analysis for CRP levels

In the 1KG meta-analysis, 8,002 variants were associated with CRP at $P<5\times10^{-8}$ (QQ-plot Figure S3; Manhattan plot Figure S4). This resulted in 40 distinct loci, of which 36 overlapped with the HapMap meta-analysis (Table 1). The lead variant at the *CRP* locus in the 1KG GWAS was rs4287174 (β=-0.185, $P=1.95\times^{-398}$), which is in high LD with rs2794520 ($r^2=0.98$), the lead variant at the *CRP* locus in the HapMap GWAS. Among eight of the overlapping loci, the lead variant was at the same position in both

GWAS (rs1260326, rs1490384, rs10832027, rs1582763, rs7310409, rs2239222, rs340005, and rs1800961). Compared with HapMap, the four additional variants identified in 1KG were rs75460349, rs1514895, rs112635299, and rs1189402. The variants rs1514895 and rs1189402 were available in the HapMap GWAS, but were not associated at the genome-wide threshold (respectively $P=1.2\times10^{-7}$ and $P=8.1\times10^{-3}$). The two variants rs75460349 and rs112635299 were not available in the HapMap GWAS, nor were variants in high LD ($r^2<0.8$). The rs75460349 is a low frequency variant with a coded allele frequency of 0.97 ($\beta=0.086$, $P=4.5\times10^{-10}$). Also rs112635299 near the *SERPINA1/2* gene is a low frequency variant with a MAF of 0.02 ($\beta=0.107$, $P=2.1\times10^{-10}$). Adjustment for BMI in the 1KG GWAS (n=147,827) revealed six additional loci that were not associated with CRP in the HapMap and 1KG primary analyses (Table 1; Table S5, QQ-plot Figure S5; Manhattan plot Figure S6). The associations at three lead variants were much reduced after adjustment for BMI (rs1558902 (*FTO*), $P_{adjusted}=0.40$; rs12995480 (*TMEM18*), $P_{adjusted}=0.02$; rs64343 (*ABO*), $P_{adjusted}=1.0\times10^{-7}$). Both the *FTO* and *TMEM18* gene are well-known obesity genes. Except for the *FTO*, *TMEM18*, and *ABO* loci, all distinct loci identified in the primary 1KG analysis were also associated with CRP in the BMI adjusted 1KG analysis. No genome-wide significant association was observed on the X-chromosome in the 1KG GWAS including 102,086 individuals.

LD score regression

The HapMap GWAS LD Score regression intercept was 1.03 (standard error: 0.013), and the 1KG intercept was 1.02 (standard error 0.011). This suggests that a small proportion of the inflation is attributable to confounding bias (~12% for the HapMap GWAS and ~13% for the 1KG GWAS). Hence, the vast majority of inflation is due to the polygenic architecture of circulating CRP levels. As depicted in Figure 1, CRP showed strong positive genetic correlations with anthropometric traits (e.g., BMI: $R_g=0.43$, $P=5.4\times10^{-15}$), glycemic phenotypes (e.g., type 2 diabetes $R_g=0.33$, $P=3.1\times10^{-6}$), lipid phenotypes (e.g., triglycerides $R_g=0.29$, $P=7.9\times10^{-5}$), and coronary artery disease ($R_g=0.23$, $P=2.4\times10^{-5}$) (Table S6). By comparison, CRP showed inverse genetic correlations with educational attainment (e.g., college

completion R_g=-0.27, P=9.2×10^{-7}), lung function (e.g., forced vital capacity R_g=-0.24, P=4.6×10^{-12}), and HDL-cholesterol (R_g=-0.30, P=4.8×10^{-9}).

Additional signals at distinct loci

Approximate conditional analyses in the 1KG GWAS revealed additional signals at nine loci (Table S7). Five loci showed one secondary signal (*IL6R, NLRP3, HNF1A, CD300LF,* and *APOE/APOC1*), the *PPP1R3B* locus had two additional signals, the *LEPR* locus had three additional signals, and the *SALL1* locus had four additional signals, whereas the *CRP* locus showed a total of 13 distinct associated variants. Interestingly, the rs149520992 rare variant (MAF=0.01) mapping to the *CRP* locus showed an association at $P_{conditional}=3.7\times10^{-15}$ with β=-0.272 for the T-allele. The GCTA effect estimates for the ten distinct variants in the vicinity of the *CRP* gene identified in the 1KG conditional analysis are in high correlation with the effect estimates of these variants obtained from the RS-I and WGHS individual level data (r_{RS}=0.97, and r_{WGHS}=0.84), confirming the reliability of the GCTA estimates.

Variance explained of CRP

The lead variant at the *CRP* locus in both the HapMap (rs2794520) and 1KG (rs4287174) GWAS explained 1.4% of the variance in natural log-transformed CRP levels. The distinct variants at the *CRP* locus derived from the joint conditional analysis in the 1KG GWAS explained 4.3% of the variance. The lead variants at all distinct loci together explained 6.2% of the CRP variance in the HapMap GWAS, and 6.5% in the 1KG GWAS. When we added the distinct variants at associated loci derived from the conditional analysis, the variance explained by all associated loci was 11.0% in the 1KG GWAS.

Functional annotation

We applied DEPICT and MAGMA analyses for functional annotation and biological interpretation of the findings. The DEPICT analysis included 9,497 genome-wide significant variants, covering 283 genes, and prioritized 55 candidate genes across 29 regions (FDR<0.05, Table S8). The prioritized genes included

IL6R mapping to the 1q21.3 locus (represented by rs4129267) and *APCS* to the 1q32.2 locus. Investigating 10,968 reconstituted gene sets for enrichment, DEPICT highlighted 583 (5.3%) gene sets to be significantly enriched among CRP-associated loci at FDR<0.05 (Table S9). Using further clustering, we identified 66 groups of gene sets that substantially correlated and clustered in two sets, one mainly comprised of immune pathways, and the other enriched for metabolic pathways (Figure 2). In Figure 3, we present the prioritized genes and the most significant gene sets. We found synovial fluid, liver tissue, and monocytes to be enriched for the expression of the prioritized genes (FDR<0.05). The MAGMA analysis was applied on the HapMap GWAS, identifying five significantly enriched gene sets (Bonferroni-corrected P<0.05, Table S10). Results included consequences of gene EGF induction, positive regulation of gene expression, and IL-6 signaling pathway, in line with the most strongly prioritized gene from DEPICT gene prioritization. MAGMA analysis prioritized liver as a sole enriched tissue (P=0.048).

To prioritize the most likely trait-relevant gene for each GWAS locus, we interrogated the GWAS data with *cis*-eQTL data identified from 44 post-mortem tissues and a large whole blood eQTL meta-analysis using colocalization analysis (Table S11). Figure S7 presents the GWAS loci that colocalize with *cis*-eQTLs with the corresponding tissue, the colocalizing gene, and the posterior probability of one shared underlying variant driving both associations. Out of the 58 lead gSNPs, 25 SNPs (43%) showed evidence of colocalization with one or more local eQTL effects (posterior probability >0.9). For example, the rs2293476 locus colocalizes with several *cis*-eQTL effects for *PABC4*, and pseudogenes *OXCT2P1*, *RP11−69E11.4*, and *RP11−69E11.8*. The rs10925027 locus shows colocalization with cis-eQTL effect for *NLRP3*, exclusively in the highly powered blood meta-analysis. Out of 25 loci, for nine loci there was only one colocalizing gene. Altogether, gSNP-associated cis-eQTL effects were present in up to 14 different tissues, with whole blood, esophagus mucosa, skin, and tibial nerve being the most frequent.

Mendelian randomization analyses

We observed a protective effect of genetically determined variance in CRP with schizophrenia with an IVW odds ratio (OR) of the 52-SNP score of 0.89 (95%CI: 0.81-0.97, $P=6.6\times10^{-3}$) (Tables S14-S15,

Figure S8-S11). The MR-Egger intercept was compatible with no unbalanced pleiotropy (P=0.48). The estimate of the rs2794520 variant was comparable to the 52-SNP score estimate (OR 0.89, 95% CI 0.84-0.94, P=0.046). The WM and PWM estimates were comparable to the IVW estimate (OR_{WM} 0.89, $P_{WM}=5.1\times10^{-3}$; OR_{PWM} 0.89, $P_{PWM}=4.4\times10^{-3}$). The "leave-one-out" analysis provided evidence that no single variant was driving the IVW point estimate (Figure S10). The causal OR between the rs2794520 variant and BD was 1.33 (95% CI 1.03-1.73, P=0.032). For the 52-SNP score, the IVW OR was 1.16 (95% CI 1.00-1.35, P=0.054). The MR-Egger intercept was compatible with unbalanced pleiotropy (P=0.049). The MR-Egger estimate OR of the 52-SNP score was comparable to the rs2794520 estimate (OR 1.36, 95%CI 1.1.-1.69, $P=6.7\times10^{-3}$), as were the WM and PWM estimates (OR_{WM} 1.33, $P_{WM}=3.4\times10^{-3}$; OR_{PWM} 1.32, $P_{PWM}=4.3\times10^{-3}$).

We observed evidence against a causal association between either *CRP* rs2794520 (OR 1.01, 95%CI 0.91-1.12, P=0.88), or the 52-SNP instrument (OR 0.96, 95%CI 0.84-1.09, P=0.51) and CAD. An Egger intercept of 0.014 suggested presence of unbalanced pleiotropy ($P=5.8\times10^{-3}$), with an MR-Egger causal estimate of OR 0.79 (95%CI 0.67-0.94, P=0.012). However, the WM and PWM showed no association between CRP and CAD. For AD, there was evidence against an association with rs2794520 (P=0.592), though the IVW OR showed a protective effect (OR 0.51, 95%CI 0.30-0.88, P=0.015). The Egger intercept of 0.046 suggested unbalanced pleiotropy (P=0.042), and the MR-Egger OR was 0.27 (95%CI 0.12-0.60). However, the association was null for the WM and PWM analyses (OR_{WM} 1.04, $P_{WM}=0.61$; OR_{PWM} 1.05, $P_{PWM}=0.53$). We observed evidence against an effect for CD, DBP, IBD, RA, and SBP for the rs2794520 variant and the IVW, MR-Egger, WM, and PWM analyses.

Discussion

Using genomic data from >200,000 individuals, we have identified 58 distinct signals for circulating CRP levels, confirming 16 previously identified CRP loci. BMI-adjusted GWAS suggested that the vast majority of genetic risk variants affect CRP levels independent of its main determinant (BMI). The genome-wide *in silico* functional annotation analysis highlights 55 genes which are likely to explain the

association of 29 signals to CRP levels. The data identified gene sets involved in the biology of immune system and liver as main regulators of serum CRP levels. Mendelian randomization analyses supported causal associations of genetically increased CRP with a protective effect on schizophrenia, and increased risk of bipolar disorder.

Obesity is one of the main determinants of chronic low-grade inflammation in the general population[37; 38]. Adjustment for BMI in the CRP GWAS abolished the association at only three lead variants, suggesting that the genetic regulation of chronic low-grade inflammation is largely independent from BMI. Notably, BMI adjustment resulted in the identification of six variants that were not associated with CRP in the BMI-unadjusted GWAS. This supports the notion that adjustment for covariates that explain phenotypic variance may improve the statistical power in linear model analyses of quantitative traits[39]. Although adjustment for heritable correlated traits in GWAS may bias effect estimates (collider bias)[40], there is consistent evidence in the literature that BMI has a causal direct effect on CRP levels[41], and therefore, collider bias in CRP GWAS adjusted for BMI is less likely.

The sex-stratified analyses revealed significant heterogeneity in effect estimates between men and women at only four lead variants, which represent less than 10% of all CRP loci. Even among these four loci the effect directions were similar, thus the heterogeneity was limited to effect sizes. The data suggest that the difference between men and women in CRP levels is less likely to be explained by genetic factors. Furthermore, two signals identified in the former HapMap GWAS of CRP levels were not significant in the current HapMap GWAS. The effect estimates in the current analyses were too small to identify with our sample size.

The top variant at the *CRP* locus in both the HapMap and 1KG GWAS explained 1.4% of the variance in circulating CRP levels. The approximate conditional analysis resulted in 13 variants jointly associated within the *CRP* locus in the 1KG GWAS. With respect to locus definition, we used a conservative distance criterion compared to other GWA studies that often use ±500kb surrounding the GWAS peak[42]. Here, we used the criterion that the minimum distance between the boundaries of loci is 500kb. In order to identify further variants associated with CRP levels, we performed approximate

conditional analyses resulting in multiple putative additional variants, also inside and near genes that were not identified in the primary GWAS. As an example, the *CRP* locus spanned >2MB according to our criterion. Approximate conditional analysis revealed that two variants, namely rs3027012 near *DARC* and rs56288844 near *FCER1A*, both downstream of the *CRP* gene, were associated with CRP levels. Furthermore, upstream of *CRP*, we identified a variant near *FCGR2A* (Immunoglobulin G Fc Receptor II). These results show that for a given lead variant, potentially multiple causal loci, here *DARC*, *FCER1A*, and *FCGRA2,* alongside *CRP* contribute to chronic low-grade inflammation and variation in circulating CRP levels.

DEPICT analysis provided further evidence that the genes annotated to the associated CRP variants mainly cluster in the immune and liver biological systems. Notably, the gene set "inflammatory response", which captures both immune response and liver metabolism, was the main connector network between the two networks. This is in line with the observation that CRP is mainly produced by liver cells in response to inflammatory cytokines during acute and chronic inflammation[43]. Interestingly, the analysis highlighted iron homeostasis as an enriched gene set. In agreement, the conditional analysis highlighted a distinct genetic association at the hemochromatosis gene *HFE,* a transmembrane protein of the major histocompatibility complex (MHC) class I family. Previous studies show that iron metabolism plays a pivotal role in inflammation[44; 45]. However, genetic pleiotropy may highlight co-regulated networks in pathway analysis that are not causal to inflammation per se. It is also important to note that the results of DEPICT analyses apply to reconstituted gene sets which may sometimes have slightly different overlaying biological theme than the original gene set annotation.

The MR analyses validate previous evidence that genetically-elevated CRP is protective for the risk of schizophrenia[13; 46], although observational data suggest a positive association between CRP and risk of schizophrenia[47]. For bipolar disorder we observed a positive causal effect, which is in line with previous MR and observational studies[13; 48]. Although the causal underlying mechanisms remain to be elucidated, a hypothesis for the schizophrenia observation might be the immune response to infections early in life. Levels of acute-phase response proteins in dry blood spots collected at birth are lower for

patients with non-affective psychosis, which includes schizophrenia, compared to controls, suggesting a weaker immune response at birth[49]. Also, neonates that have been exposed to a maternal infection and have low levels of acute-phase response proteins, have a higher risk of schizophrenia[50]. Altogether, the evidence suggests that a deficient immune response may contribute to chronic infection in children and the development of schizophrenia. For AD and CAD, the Egger intercept showed evidence of unbalanced pleiotropy and the Egger estimate showed a protective effect of CRP on the risk of AD and CAD. However, for both outcomes, the effects of the WM and PWM analyses, as well as analyses using the single rs2794520 variant (which is least likely to be affected by pleiotropy) were null. The MR-Egger estimate relies on the InSIDE assumption which states that the strength of the association between the genetic variants and CRP is independent from the strength of the direct pleiotropic effects of the genetic variants on the outcome. This assumption may be violated when the genetic variants are associated with a confounder of the CRP-outcome association. Such a scenario may occur when the genetic variants are associated with an exposure that is causally upstream of the exposure under study. In the context of the association of CRP with AD and CAD, this could be lipids or glycemic phenotypes. Several genetic variants used in the 52-SNP instrument are associated with metabolic phenotypes that may affect CRP levels. In agreement, the WM and PWM, in which the InSIDE assumption is relaxed, and the single variant analysis showed no association. Furthermore, the observation that CRP is not causally related to CAD in the MR analyses is in comparison to previous published studies[51]. Power calculation showed that we had 100% power to detect a 10% difference in CAD risk, thus the probability of a false negative finding is small. Also, CRP is associated with future CAD in observational studies, and randomized trials have shown a beneficial effect of lowering inflammation using statins[52] and canakinumab[53] on CAD risk, but this effect is unlikely to be attributable to CRP.

The strengths of our study are the use of a very large sample size for CRP and the use of both HapMap and 1KG imputed data. Furthermore, we conducted sex-specific and BMI-adjusted analyses to study the effect of sex and body mass on the associations between genetic variants and CRP. To maximize power and to efficiently use the data, we meta-analyzed all available samples in a discovery setting

without replication. The consistent association of the variants in >50 studies at a strict Bonferroni corrected threshold provide confidence that our findings represent true associations. We used both HapMap and 1KG imputed data to identify genetic variants for circulating CRP levels. At the start of the project, more studies had HapMap imputed data available. Hence, the sample size and thus power in the HapMap GWAS was higher compared to the 1KG. Also, HapMap may identify variants that are not identified in 1KG GWAS[54]. Nevertheless, 1KG offers better coverage of uncommon variants and includes INDELs, which are not included in the HapMap reference panel. Including both reference panels, we used all available samples and maximized the possibility to identify genetic variants for CRP, both common and uncommon.

However, we note limitations to our study. GWAS merely identify loci associated with complex phenotypes and the identification of causal genes remains challenging. We only included individuals of European ancestry; the generalizability of our findings to other races/ethnicities is uncertain. In addition, although our analyses provided support for causal associations, we acknowledge that we may not have identified the causal variants and we may not have eliminated residual confounding. The colocalisation analyses provide evidence for colocalisation of CRP GWAS signals and eQTLs, however, it does not provide evidence that the GWAS signal is functioning on CRP through the gene expression. We further note that the method assumes identical LD-structure from the GWAS and eQTL datasets. As there are ~14% of non-European samples in the full GTEx dataset, this assumption might be violated for some tissues. Last, we meta-analyzed all available samples in one meta-analysis and did not replicate our findings in an independent sample. Therefore, our findings may need replication.

In conclusion, we performed a large GWAS meta-analysis to identify genetic loci associated with circulating CRP levels, a sensitive marker of chronic low-grade inflammation, and found support for a causal role of CRP with a decreased risk of schizophrenia and higher risk of bipolar disorder. As inflammation is implicated in the pathogenesis of multiple complex diseases, the new insights into the biology of inflammation obtained in the current study may contribute to future therapies and interventions.

Supplemental data

Supplemental data include 11 figures, 17 tables, and the study-specific descriptives and acknowledgments.

Acknowledgements

The participating studies report the acknowledgments in the study-specific supplemental information.

Declaration of interest

OHF works in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.), Metagenics Inc., and AXA. Nestlé Nutrition (Nestec Ltd.), Metagenics Inc., and AXA had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review or approval of the manuscript. Other authors report no conflict of interest.

Web Resources

DEPICT, https://data.broadinstitute.org/mpg/depict/

GCTA, cnsgenomics.com/software/gcta/

GIANT 1KG p1v3 EUR reference panel, http://csg.sph.umich.edu/abecasis/mach/download/1000G.2012-03-14.html

GTEx Portal, https://www.gtexportal.org/

GWAMA, https://www.geenivaramu.ee/en/tools/gwama

LD Hub, ldsc.broadinstitute.org/ldhub/

METAL, https://genome.sph.umich.edu/wiki/METAL_Documentation

mRnd power calculations for Mendelian Randomization, http://cnsgenomics.com/shiny/mRnd/

Summary data

The full GWAS results of the HapMap and 1KG meta-analyses can be found at <dbGAP link to be added>. Codes used for the downstream analyses can be found at <GitHub link to be added>.

References

1. Libby, P. (2012). Inflammation in atherosclerosis. Arterioscler Thromb Vasc Biol 32, 2045-2051.
2. Pickup, J.C. (2004). Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. Diabetes Care 27, 813-823.
3. Akiyama, H., Barger, S., Barnum, S., Bradt, B., Bauer, J., Cole, G.M., Cooper, N.R., Eikelenboom, P., Emmerling, M., Fiebich, B.L., et al. (2000). Inflammation and Alzheimer's disease. Neurobiol Aging 21, 383-421.
4. Khandaker, G.M., Cousins, L., Deakin, J., Lennox, B.R., Yolken, R., and Jones, P.B. (2015). Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment. Lancet Psychiatry 2, 258-270.
5. Pepys, M.B. (1995). The acute phase response and C-reactive protein. Oxford textbook of medicine 2, 1527-1533.
6. Danesh, J., Wheeler, J.G., Hirschfield, G.M., Eda, S., Eiriksdottir, G., Rumley, A., Lowe, G.D., Pepys, M.B., and Gudnason, V. (2004). C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N Engl J Med 350, 1387-1397.
7. Dehghan, A., Kardys, I., de Maat, M.P., Uitterlinden, A.G., Sijbrands, E.J., Bootsma, A.H., Stijnen, T., Hofman, A., Schram, M.T., and Witteman, J.C. (2007). Genetic variation, C-reactive protein levels, and incidence of diabetes. Diabetes 56, 872-878.
8. Wium-Andersen, M.K., Ørsted, D.D., and Nordestgaard, B.G. (2013). Elevated C-reactive protein associated with late-and very-late-onset schizophrenia in the general population: a prospective study. Schizophr Bull, sbt120.
9. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P., et al. (2011). Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation 123, 731-738.
10. de Vries, P.S., Sabater-Lleal, M., Chasman, D.I., Trompet, S., Ahluwalia, T.S., Teumer, A., Kleber, M.E., Chen, M.H., Wang, J.J., Attia, J.R., et al. (2017). Comparison of HapMap and 1000 Genomes Reference Panels in a Large-Scale Genome-Wide Association Study. PLoS One 12, e0167742.
11. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101, 5-22.
12. Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N., and Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 27, 1133-1163.
13. Prins, B.P., Abbasi, A., Wong, A., Vaez, A., Nolte, I., Franceschini, N., Stuart, P.E., Guterriez Achury, J., Mistry, V., Bradfield, J.P., et al. (2016). Investigating the causal relationship of C-reactive protein with 32 complex somatic and psychiatric outcomes: a large-scale cross-consortium Mendelian randomization study. PLoS Med 13, e1001976.
14. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C.M., and Boerwinkle, E. (2009). Cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2, 73-80.
15. Mägi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 11, 288.
16. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190-2191.
17. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpelainen, T.O., Esko, T., Magi, R., Li, S., et al. (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. PLoS Genet 9, e1003500.

18. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291-295.
19. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S., et al. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33, 272-279.
20. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236-1241.
21. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76-82.
22. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44, 369-375.
23. Park, J.-H., Wacholder, S., Gail, M.H., Peters, U., Jacobs, K.B., Chanock, S.J., and Chatterjee, N. (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42, 570-575.
24. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 6, 5890.
25. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.
26. Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. Bioinformatics 27, 2463-2464.
27. Kolde, R. (2012). Pheatmap: pretty heatmaps. R package version 61.
28. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 11, e1004219.
29. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat commun 8, 1826.
30. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10, e1004383.
31. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 45, 1238-1243.
32. GTEx Consortium. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348, 648-660.
33. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. eLife 7, e34408.
34. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 44, 512-525.
35. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol 40, 304-314.
36. Brion, M.-J.A., Shakhbazov, K., and Visscher, P.M. (2012). Calculating statistical power in Mendelian randomization studies. Int J Epidemiol 42, 1497-1501.

37. Visser, M., Bouter, L.M., McQuillan, G.M., Wener, M.H., and Harris, T.B. (1999). Elevated C-reactive protein levels in overweight and obese adults. JAMA 282, 2131-2135.
38. Wellen, K.E., and Hotamisligil, G.S. (2003). Obesity-induced inflammatory changes in adipose tissue. J Clin Invest 112, 1785.
39. Robinson, L.D., and Jewell, N.P. (1991). Some surprising results about covariate adjustment in logistic regression models. International Statistical Review/Revue Internationale de Statistique, 227-240.
40. Aschard, H., Vilhjálmsson, B.J., Joshi, A.D., Price, A.L., and Kraft, P. (2015). Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. Am J Hum Genet 96, 329-339.
41. Timpson, N.J., Nordestgaard, B.G., Harbord, R.M., Zacho, J., Frayling, T.M., Tybjærg-Hansen, A., and Smith, G.D. (2011). C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. Int J Obes 35, 300-308.
42. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197-206.
43. Moshage, H.J., Roelofs, H.M.J., Van Pelt, J.F., Hazenberg, B.P.C., Van Leeuwen, M.A., Limburg, P.C., Aarden, L.A., and Yap, S.H. (1988). The effect of interleukin-1, interleukin-6 and its interrelationship on the synthesis of serum amyloid A and C-reactive protein in primary cultures of adult human hepatocytes. Biochem Biophys Res Commun 155, 112-117.
44. Ganz, T., and Nemeth, E. (2015). Iron homeostasis in host defence and inflammation. Nat Rev Immunol 15, 500-510.
45. Alizadeh, B., Njajou, O., Hazes, J., Hofman, A., Slagboom, P., Pols, H., and van Duijn, C. (2007). The H63D variant in the HFE gene predisposes to arthralgia, chondrocalcinosis and osteoarthritis. Ann Rheum Dis 66, 1436-1442.
46. Hartwig, F.P., Borges, M.C., Horta, B.L., Bowden, J., and Smith, G.D. (2017). Inflammatory Biomarkers and Risk of Schizophrenia: A 2-Sample Mendelian Randomization Study. JAMA psychiatry.
47. Fernandes, B.S., Steiner, J., Bernstein, H.-G., Dodd, S., Pasco, J., Dean, O., Nardin, P., Goncalves, C.-A., and Berk, M. (2016). C-reactive protein is increased in schizophrenia but is not altered by antipsychotics: meta-analysis and implications. Mol Psychiatry 21, 554-565.
48. Fernandes, B.S., Steiner, J., Molendijk, M.L., Dodd, S., Nardin, P., Gonçalves, C.-A., Jacka, F., Köhler, C.A., Karmakar, C., Carvalho, A.F., et al. (2016). C-reactive protein concentrations across the mood spectrum in bipolar disorder: a systematic review and meta-analysis. Lancet Psychiatry 3, 1147-1156.
49. Gardner, R., Dalman, C., Wicks, S., Lee, B., and Karlsson, H. (2013). Neonatal levels of acute phase proteins and later risk of non-affective psychosis. Transl psychiatry 3, e228.
50. Blomström, Å., Gardner, R., Dalman, C., Yolken, R., and Karlsson, H. (2015). Influence of maternal infections on neonatal acute phase proteins and their interaction in the development of non-affective psychosis. Transl psychiatry 5, e502.
51. Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C., and Bennett, D. (2009). Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. JAMA 302, 37-48.
52. Ridker, P.M., Danielson, E., Fonseca, F.A., Genest, J., Gotto, A.M., Jr., Kastelein, J.J., Koenig, W., Libby, P., Lorenzatti, A.J., MacFadyen, J.G., et al. (2008). Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 359, 2195-2207.
53. Ridker, P.M., Everett, B.M., Thuren, T., MacFadyen, J.G., Chang, W.H., Ballantyne, C., Fonseca, F., Nicolau, J., Koenig, W., Anker, S.D., et al. (2017). Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N Engl J Med 377, 1119-1131.
54. Wood, A.R., Perry, J.R., Tanaka, T., Hernandez, D.G., Zheng, H.-F., Melzer, D., Gibbs, J.R., Nalls, M.A., Weedon, M.N., Spector, T.D., et al. (2013). Imputation of variants from the 1000 Genomes

Project modestly improves known associations and can identify low-frequency variant-phenotype associations undetected by HapMap based imputation. PLoS One 8, e64343.

Figure Legends

Figure 1. Genome-wide genetic correlation between serum CRP levels and different phenotypes and clinical diseases. The genetic correlation and its standard error are estimated with linkage disequilibrium score regression analysis. ADHD, attention deficit and hyperactivity disorder; FEV1, forced expiratory volume in 1 second; FVC, forced vital capacity; HOMA-B, homeostatic model assessment β-cell function; HOMA-IR, homeostatic model assessment insulin resistance; HbA1C, Hemoglobin A1c.

Figure 2. Results of the DEPICT functional annotation analysis. Each node represents exemplar gene set from Affinity Propagation clustering and links represent corresponding Pearson correlation coefficients between individual enriched gene sets (only the links with r>0.3 are shown). As an example, outlined are the individual gene sets inside two clusters ("Inflammatory response" and "negative regulation of peptidase activity").

Figure 3. Heatmap representing the results of DEPICT functional annotation analysis. Each row represents enriched (FDR < 0.05) gene sets and each column represents prioritized (FDR<0.05) genes. Colors on the heatmap represent each gene´s contribution to gene set enrichment (depicted as Z-score, only top 10 highest Z-scores per gene set are visualized). Sidebars represent p-values for GWAS, gene set enrichment (GSE), and gene prioritization (nominal P-value on $\log_{10}$ scale). Top 10 gene sets per annotation category are visualized. GO, Gene Ontology; KE, Kyoto Encyclopedia of Gene and Genomes; RE, REACTOME pathways; MP, Mouse Phenotypes; PI, protein-protein interactions.

24 **Table 1. Newly identified loci associated with C-reactive protein.**

Variant	Chr	Position	Coded allele	Coded Allele freq	Beta	SE	P-value	Closest Gene	1KG lead variant
Loci found in the HapMap GWAS									
rs469772	1	91530305	T	0.19	-0.031	0.005	5.54×10^{-12}	*ZNF644*	rs469882
rs12995480	2	629881	T	0.17	-0.031	0.005	1.24×10^{-10}	*TMEM18*	rs62105327
rs4246598	2	88438050	A	0.46	0.022	0.004	5.11×10^{-10}	*FABP1*	-
rs9284725	2	102744854	C	0.24	0.027	0.004	7.34×10^{-11}	*IL1R1*	rs1115282
rs1441169	2	214033530	G	0.53	-0.025	0.004	2.27×10^{-11}	*IKZF2*	-
rs2352975	3	49891885	C	0.30	0.025	0.004	6.43×10^{-10}	*TRAIP*	rs10049413
rs17658229	5	172191052	C	0.05	0.056	0.010	5.50×10^{-09}	*DUSP1*	rs34471628
rs9271608	6	32591588	G	0.22	0.042	0.005	2.33×10^{-17}	*HLA-DQA1*	rs2647062
rs12202641	6	116314634	T	0.39	-0.023	0.004	3.00×10^{-10}	*FRK*	-
rs1490384	6	126851160	T	0.51	-0.025	0.004	2.65×10^{-12}	*C6orf173*	rs1490384
rs9385532	6	130371227	T	0.33	-0.026	0.004	1.90×10^{-11}	*L3MBTL3*	-
rs1880241	7	22759469	G	0.48	-0.028	0.004	8.41×10^{-14}	*IL6*	rs13241897
rs2710804	7	36084529	C	0.37	0.021	0.004	1.30×10^{-08}	*KIAA1706*	-
rs2064009	8	117007850	C	0.42	-0.027	0.004	2.28×10^{-14}	*TRPS1*	rs6987444
rs2891677	8	126344208	C	0.46	-0.020	0.004	1.59×10^{-08}	*NSMCE2*	rs10956251
rs643434	9	136142355	A	0.37	0.023	0.004	1.02×10^{-09}	*ABO*	9:136146061
rs1051338	10	91007360	G	0.31	0.024	0.004	2.27×10^{-09}	*LIPA*	-
rs10832027	11	13357183	G	0.33	-0.026	0.004	4.43×10^{-12}	*ARNTL*	rs10832027
rs10838687	11	47312892	G	0.22	-0.031	0.004	9.12×10^{-13}	*MADD*	rs7125468
rs1582763	11	60021948	A	0.37	-0.022	0.004	2.37×10^{-09}	*MS4A4A*	rs1582763
rs7121935	11	72496148	A	0.38	-0.022	0.004	5.28×10^{-09}	*STARD10*	-
rs11108056	12	95855385	G	0.42	-0.028	0.004	5.42×10^{-14}	*METAP2*	rs12813389
rs2239222	14	73011885	G	0.36	0.035	0.004	9.87×10^{-20}	*RGS6*	rs2239222
rs4774590	15	51745277	A	0.35	-0.022	0.004	2.71×10^{-08}	*DMXL2*	rs1189402
rs1558902	16	53803574	A	0.41	0.034	0.004	5.20×10^{-20}	*FTO*	rs55872725
rs178810	17	16097430	T	0.56	0.020	0.004	2.95×10^{-08}	*NCOR1*	-
rs10512597	17	72699833	T	0.18	-0.037	0.005	4.44×10^{-14}	*CD300LF,RAB37*	rs2384955
rs4092465	18	55080437	A	0.35	-0.027	0.004	3.11×10^{-10}	*ONECUT2*	-
rs12960928	18	57897803	C	0.27	0.024	0.004	1.91×10^{-09}	*MC4R*	-
rs2315008	20	62343956	T	0.31	-0.023	0.004	5.36×10^{-10}	*ZGPAT*	-
rs2836878	21	40465534	G	0.27	0.043	0.004	7.71×10^{-26}	*DSCR2*	rs4817984
rs6001193	22	39074737	G	0.35	-0.028	0.004	6.53×10^{-14}	*TOMM22*	rs4821816

Additional loci found in the 1KG GWAS								
rs75460349	1	27180088	A	0.97	0.086	0.014	4.50×10^{-10}	*ZDHHC18*
rs1514895	3	170705693	A	0.71	-0.027	0.004	2.70×10^{-09}	*EIF5A2*
rs112635299	14	94838142	T	0.02	-0.107	0.017	2.10×10^{-10}	*SERPINA1/2*
rs1189402	15	53728154	A	0.62	0.025	0.004	3.90×10^{-09}	*ONECUT1*
Additional loci found in the BMI adjusted 1KG GWAS								
3:47431869	3	47431869	D	0.59	0.024	0.004	1.10×10^{-08}	*PTPN23*
rs687339	3	135932359	T	0.78	-0.030	0.005	2.80×10^{-10}	*MSL2*
rs7795281	7	74122854	A	0.76	0.028	0.005	3.10×10^{-08}	*GTF2I*
rs1736060	8	11664738	T	0.60	0.029	0.004	2.60×10^{-13}	*FDFT1*
17:58001690	17	58001690	D	0.44	-0.026	0.004	9.50×10^{-10}	*RPS6KB1*
rs9611441	22	41339367	C	0.49	-0.022	0.004	1.40×10^{-08}	*XPNPEP3*

β coefficient represents 1-unit change in the natural log-transformed CRP (mg/L) per copy increment in the allele A1. Freq is the frequency of A1. Position is according to Hg19. When a variant is located within a gene, that gene is reported in the closest gene column, otherwise the closest gene. The HapMap variants are presented, except for the 1KG additional findings. For the HapMap loci, the lead variant from the 1KG GWAS is presented when the locus was also found in the 1KG GWAS.