

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

3
4 Symen Ligthart¹, Ahmad Vaez², Urmo Vösa³, Maria G Stathopoulou⁴, Paul S de Vries⁵, Bram P Prins⁶,
5 Peter J Van der Most², Toshiko Tanaka⁷, Elnaz Naderi^{2, 8}, Lynda M Rose⁹, Ying Wu¹⁰, Robert Karlsson¹¹,
6 Maja Barbalic¹², Honghuang Lin¹³, René Pool^{14, 15}, Gu Zhu¹⁶, Aurélien Macé¹⁷⁻¹⁹, Carlo Sidore²⁰, Stella
7 Trompet^{21, 22}, Massimo Mangino^{23, 24}, Maria Sabater-Lleal²⁵, John P Kemp^{26, 27}, Ali Abbasi^{2, 28, 29}, Tim
8 Kacprowski^{30, 31}, Niek Verweij³², Albert V. Smith^{33, 34}, Tao Huang^{35, 36}, Carola Marzi^{37, 38}, Mary F.
9 Feitosa³⁹, Kurt K Lohman⁴⁰, Marcus E Kleber⁴¹, Yuri Milaneschi⁴², Christian Mueller⁴³⁻⁴⁵, Mahmudul
10 Huq², Efthymia Vlachopoulou⁴⁶, Leo-Pekka Lyytikäinen^{47, 48}, Christopher Oldmeadow^{49, 50}, Joris Deelen^{51,}
11 ⁵², Markus Perola⁵³, Jing Hua Zhao⁵⁴, Bjarke Feenstra⁵⁵, LifeLines Cohort Study², Marzyeh Amini²,
12 CHARGE Inflammation Working Group, Jari Lahti⁵⁶⁻⁵⁸, Katharina E Schraut⁵⁹, Myriam Fornage⁶⁰,
13 Bhoom Suktitipat⁶¹, Wei-Min Chen⁶², Xiaohui Li⁶³, Teresa Nutile⁶⁴, Giovanni Malerba⁶⁵, Jian'an Luan⁵⁴,
14 Tom Bak⁶⁶, Nicholas Schork^{67, 68}, Fabiola Del Greco M⁶⁹, Elisabeth Thiering^{70, 71}, Anubha Mahajan⁷²,
15 Riccardo E Marioni^{73, 74}, Evelin Mihailov⁷⁵, Joel Eriksson⁷⁶, Ayse Bilge Ozel⁷⁷, Weihua Zhang^{78, 79}, Maria
16 Nethander⁸⁰, Yu-Ching Cheng⁸¹, Stella Aslibekyan⁸², Wei Ang⁸³, Iliaria Gandin⁸⁴, Loïc Yengo⁸⁵, Laura
17 Portas⁸⁶, Charles Kooperberg⁸⁷, Edith Hofer^{88, 89}, Kumar B Rajan⁹⁰, Claudia Schurmann^{91, 92}, Wouter den
18 Hollander⁹³, Tarunveer S Ahluwalia^{94, 95}, Jing Zhao⁹⁶, Harmen HM Draisma^{14, 15, 97}, Ian Ford⁹⁸, Nicholas
19 Timpson⁹⁹, Alexander Teumer¹⁰⁰, Hongyan Huang¹⁰¹, Simone Wahl^{37, 38}, YongMei Liu⁴⁰, Jie Huang¹⁰²,
20 Hae-Won Uh¹⁰³, Frank Geller⁵⁵, Peter K Joshi⁵⁹, Lisa R. Yanek¹⁰⁴, Elisabetta Trabetti⁶⁵, Benjamin Lehne⁷⁸,
21 Diego Vozzi¹⁰⁵, Marie Verbanck⁸⁵, Ginevra Biino¹⁰⁶, Yasaman Saba¹⁰⁷, Ingrid Meulenbelt⁹³, Jeff R
22 O'Connell⁸¹, Markku Laakso¹⁰⁸, Franco Giulianini⁹, Patrik KE Magnusson¹¹, Christie M Ballantyne^{109, 110},
23 Jouke Jan Hottenga¹⁴, Grant W Montgomery¹⁶, Fernando Rivadineira¹¹¹, Rico Rueedi^{17, 18}, Maristella
24 Steri²⁰, Karl-Heinz Herzig¹¹²⁻¹¹⁴, David J Stott¹¹⁵, Cristina Menni²³, Mattias Frånberg^{25, 116}, Beate St
25 Pourcain^{99, 117}, Stephan B Felix^{31, 118}, Tune H Pers^{55, 95}, Stephan JL Bakker³², Peter Kraft¹⁰¹, Annette
26 Peters¹¹⁹, Dhananjay Vaidya¹⁰⁴, Graciela Delgado⁴¹, Johannes H Smit⁴², Vera Großmann¹²⁰, Juha

1 Sinisalo¹²¹, Ilkka Seppälä^{47, 48}, Stephen R Williams¹²², Elizabeth G Holliday^{49, 50}, Matthijs Moed⁵¹, Claudia
2 Langenberg⁵⁴, Katri Räikkönen⁵⁷, Jingzhong Ding¹²³, Harry Campbell⁵⁹, Michele M Sale⁶², Yii-Der I
3 Chen⁶³, Alan L James^{124, 125}, Daniela Ruggiero^{64, 126}, Nicole Soranzo¹²⁷, Catharina A Hartman⁶⁶, Erin N
4 Smith¹²⁸, Gerald S Berenson¹²⁹, Christian Fuchsberger⁶⁹, Dena Hernandez¹³⁰, Carla MT Tiesler^{70, 71},
5 Vilmantas Giedraitis¹³¹, David Liewald⁷³, Krista Fischer⁷⁵, Dan Mellström⁷⁶, Anders Larsson¹³², Yunmei
6 Wang¹³³, William R Scott⁷⁸, Matthias Lorentzon^{76, 134}, John Beilby^{135, 136}, Kathleen A Ryan⁸¹, Craig E
7 Pennell⁸³, Dragana Vuckovic¹³⁷, Beverly Balkau¹³⁸, Maria Pina Concas¹⁰⁵, Reinhold Schmidt⁸⁸, Carlos F
8 Mendes de Leon¹³⁹, Erwin P Bottinger^{91, 140}, Margreet Kloppenburg^{141, 142}, Lavinia Paternoster²⁷, Michael
9 Boehnke¹⁴³, AW Musk^{124, 144}, Gonneke Willemsen¹⁴, David M Evans^{26, 99}, Pamela AF Madden¹⁴⁵, Mika
10 Kähönen^{146, 147}, Zoltán Kutalik^{18, 19}, Magdalena Zoledziewska²⁰, Ville Karhunen¹⁴⁸, Stephen B
11 Kritchevsky¹⁴⁹, Naveed Sattar¹⁵⁰, Genevieve Lachance²³, Robert Clarke¹⁵¹, Tamara B Harris¹⁵², Olli T
12 Raitakari^{153, 154}, John R Attia^{49, 50, 155}, Diana van Heemst²², Eero Kajantie¹⁵⁶⁻¹⁵⁸, Rossella Sorice⁶⁴, Giovanni
13 Gambaro¹⁵⁹, Robert A Scott⁵⁴, Andrew A Hicks⁶⁹, Luigi Ferrucci⁷, Marie Standl⁷⁰, Cecilia M Lindgren^{72,}
14 ^{160, 161}, John M Starr^{73, 162}, Magnus Karlsson¹⁶³, Lars Lind¹³², Jun Z Li⁷⁷, John C Chambers^{78, 79, 164-166},
15 Trevor A Mori⁸³, Eco JCN de Geus^{14, 15}, Andrew C Heath¹⁴⁵, Nicholas G Martin¹⁶, Juha Auvinen^{148, 167},
16 Brendan B Buckley¹⁶⁸, Anton JM de Craen²², Melanie Waldenberger^{37, 169}, Konstantin Strauch^{170, 171},
17 Thomas Meitinger¹⁷²⁻¹⁷⁴, Rodney J Scott^{49, 175}, Mark McEvoy^{49, 50}, Marian Beekman⁵¹, Cristina Bombieri⁶⁵,
18 Paul M Ridker^{9, 176}, Karen L Mohlke¹⁰, Nancy L Pedersen¹¹, Alanna C Morrison⁵, Dorret I Boomsma¹⁴,
19 John B Whitfield¹⁶, David P Strachan¹⁷⁷, Albert Hofman¹, Peter Vollenweider¹⁷⁸, Francesco Cucca²⁰,
20 Marjo-Riitta Javelin^{113, 148, 167, 179}, J Wouter Jukema^{21, 180, 181}, Tim D Spector²³, Anders Hamsten²⁵, Tanja
21 Zeller^{43, 45}, André G Uitterlinden^{1, 111}, Matthias Nauck^{31, 182}, Vilmundur Gudnason^{33, 34}, Lu Qi^{36, 183}, Harald
22 Grallert^{37, 38}, Ingrid B Borecki¹⁸⁴, Jerome I Rotter¹⁸⁵, Winfried März^{41, 186, 187}, Philipp S Wild^{45, 120, 188},
23 Marja-Liisa Lokki⁴⁶, Michael Boyle¹⁵⁵, Veikko Salomaa⁵³, Mads Melbye^{55, 189, 190}, Johan G Eriksson^{53, 58,}
24 ¹⁹¹, James F Wilson^{59, 192}, Brenda WJH Penninx⁴², Diane M Becker¹⁰⁴, Bradford B Worrall¹⁹³, Greg
25 Gibson⁹⁶, Ronald M Krauss¹⁹⁴, Marina Ciullo^{64, 126}, Gianluigi Zaza¹⁹⁵, Nicholas J Wareham⁵⁴, Albertine J
26 Oldehinkel⁶⁶, Lyle J Palmer¹⁹⁶, Sarah S Murray¹⁹⁷, Peter P Pramstaller^{69, 198, 199}, Stefania Bandinelli²⁰⁰,

1 Joachim Heinrich⁷⁰, Erik Ingelsson^{201, 202}, Ian J Deary^{73, 203}, Reedik Magi⁷⁵, Liesbeth Vandenput⁷⁶, Pim van
2 der Harst³², Karl C Desch²⁰⁴, Jaspal S Kooner^{79, 164, 166, 205}, Claes Ohlsson⁷⁶, Caroline Hayward¹⁹², Terho
3 Lehtimäki^{47, 48}, Alan R Shuldiner⁸¹, Donna K Arnett²⁰⁶, Lawrence J Beilin⁸³, Antonietta Robino¹⁰⁵,
4 Philippe Froguel^{85, 207}, Mario Pirastu⁸⁶, Tine Jess²⁰⁸, Wolfgang Koenig^{169, 209, 210}, Ruth JF Loos^{91, 92, 211},
5 Denis A Evans⁹⁰, Helena Schmidt^{107, 212}, George Davey Smith⁹⁹, P Eline Slagboom⁵¹, Gudny Eiríksdóttir³³,
6 Andrew P Morris^{72, 213}, Bruce M Psaty²¹⁴⁻²¹⁷, Russell P Tracy²¹⁸, Ilja M Nolte², Eric Boerwinkle^{219, 220},
7 Sophie Visvikis-Siest⁴, Alex P Reiner²¹⁵, Myron Gross²²¹, Joshua C Bis²¹⁴, Lude Franke³, Oscar H
8 Franco¹, Emilia J Benjamin^{13, 222}, Daniel I Chasman^{9, 176}, Josée Dupuis^{222, 223}, Harold Snieder², Abbas
9 Dehghan^{179*}, Behrooz Z Alizadeh^{2*}
10 Anton JM de Crean deceased.

11

12 Affiliations

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

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

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

- 1 94. Steno Diabetes Center Copenhagen, Gentofte 2820, Denmark.
- 2 95. Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of
- 3 Health and Medical Sciences, University of Copenhagen, Copenhagen 2100, Denmark.
- 4 96. Center for Integrative Genomics, School of Biology, Georgia Institute of Technology, Atlanta, GA 30332,
- 5 USA.
- 6 97. Neuroscience Campus Amsterdam, Amsterdam 1081 HV, the Netherlands.
- 7 98. Robertson Center for Biostatistics, University of Glasgow, Glasgow G12 8QQ, UK.
- 8 99. MRC Integrative Epidemiology Unit, University of Bristol, UK.
- 9 100. Department SHIP-KEF, Institute for Community Medicine, University Medicine Greifswald, Greifswald
- 10 17475, Germany.
- 11 101. Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health,
- 12 Boston, MA 2115, USA.
- 13 102. Boston VA Research Institute, Inc., Boston, MA 02130, USA.
- 14 103. Medical statistics and bioinformatics, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands.
- 15 104. GeneSTAR Research Center, Department of Medicine, Johns Hopkins University School of Medicine,
- 16 Baltimore, MD 21287, USA.
- 17 105. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste 34140, Italy.
- 18 106. Institute of Molecular Genetics, National Research Council of Italy (CNR), Pavia 27100, Italy.
- 19 107. Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of
- 20 Graz, Graz 8010, Austria.
- 21 108. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University
- 22 Hospital, Kuopio 70210, Finland.
- 23 109. Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA.
- 24 110. Methodist DeBakey Heart and Vascular Center, Houston, TX 77030, USA.
- 25 111. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam 3015 CN, the
- 26 Netherlands.
- 27 112. Department of Physiology, Institute of Biomedicine, University of Oulu, Medical Research Center Oulu and
- 28 Oulu University Hospital, Oulu 90014, Finland.
- 29 113. Biocenter Oulu, University of Oulu, Oulu 90220, Finland.
- 30 114. Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan 60-512,
- 31 Poland.
- 32 115. Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, Glasgow
- 33 G12 8QQ, UK.
- 34 116. Department of Numerical Analysis and Computer Science, Stockholm University, Stockholm 100 44,
- 35 Sweden.
- 36 117. Donders Institute, Radboud University, Nijmegen 6525 XD, the Netherlands.
- 37 118. Department for Internal Medicine B, University Medicine Greifswald, Greifswald 17475, Germany.
- 38 119. Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental
- 39 Health, Neuherber 85764, Germany.
- 40 120. Center for Thrombosis and Hemostasis (CTH), University Medical Center of the Johannes Gutenberg
- 41 University Mainz, Mainz 55131, Germany.
- 42 121. Heart and Lung Center, Helsinki University Hospital and Helsinki University, Helsinki 00029, Finland.
- 43 122. Department of Neurology, University of Virginia, Charlottesville, VA 22908, USA.
- 44 123. Department of Internal Medicine/Geriatrics, Wake Forest University Health Sciences, Winston-Salem, NC
- 45 27157, USA.
- 46 124. Busselton Population Medical Research Institute, Sir Charles Gairdner Hospital, Nedlands, WA 6009,
- 47 Australia.
- 48 125. Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, WA
- 49 6009, Australia.
- 50 126. IRCCS Neuromed, Pozzilli (IS) 86077, Italy.
- 51 127. Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK.
- 52 128. Department of Pediatrics and Rady Children's Hospital, University of California at San Diego, School of
- 53 Medicine, La Jolla, CA 92037, USA.
- 54 129. Center for Cardiovascular Health, Tulane University, New Orleans, LA 70112, USA.
- 55 130. Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA.

- 1 131. Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala 752
2 37, Sweden.
- 3 132. Department of Medical Sciences, Uppsala University, Uppsala 751 41, Sweden.
- 4 133. Department of Medicine, Case Cardiovascular Research Institute, Case Western Reserve University;
5 Harrington Heart & Vascular Institute, University Hospitals, Cleveland, OH 44106, USA.
- 6 134. Geriatric Medicine, Sahlgrenska University Hospital, Mölndal 431 80, Sweden.
- 7 135. PathWest Laboratory Medicine WA, Nedlands, WA 6009, Australia.
- 8 136. School of Pathology and Laboratory, The University of Western Australia, Crawley, Perth, WA 6009,
9 Australia.
- 10 137. Medical Sciences, Surgical and Health Department, University of Trieste, Trieste 34137, Italy.
- 11 138. Inserm U-1018, CESP, Team 5 (EpReC, Renal and cardiovascular Epidemiology), UVSQ-UPS, Villejuif,
12 94807, France.
- 13 139. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109,
14 USA.
- 15 140. Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New
16 York, NY 10029, USA
- 17 141. Department of Rheumatology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands.
- 18 142. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden 2333 ZC, the Netherlands.
- 19 143. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI
20 48109, USA.
- 21 144. Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
- 22 145. Department of Psychiatry, Washington University School of Medicine, 4560 Clayton Ave., Suite 1000, St.
23 Louis, MO, 63110, USA.
- 24 146. Department of Clinical Physiology, Tampere University Hospital, Tampere 33520, Finland.
- 25 147. Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of
26 Medicine and Life Sciences, University of Tampere, Tampere 33520, Finland.
- 27 148. Center for Life Course Epidemiology, Faculty of Medicine, University of Oulu, Oulu 90014, Finland.
- 28 149. Gerontology and Geriatric Medicine, Sticht Center on Aging and Rehabilitation, Wake Forest University
29 Health Sciences, Winston-Salem, NC 27157, USA.
- 30 150. BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow G12 8TA, UK.
- 31 151. Clinical Trial Service Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3
32 7LF, UK.
- 33 152. Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research
34 Program, National Institutes of Health, Bethesda, Maryland 20892, USA.
- 35 153. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520,
36 Finland.
- 37 154. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520,
38 Finland.
- 39 155. John Hunter Hospital, New Lambton Heights, NWS 2305, Australia.
- 40 156. Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki 00014, Finland.
- 41 157. Hospital for Children and Adolescents, Helsinki University Central Hospital and University of Helsinki,
42 Helsinki 00290, Finland.
- 43 158. Department of Obstetrics and Gynaecology, MRC Oulu, Oulu University Hospital and University of Oulu,
44 Oulu 90014, Finland.
- 45 159. Division of Nephrology and Dialysis, Columbus-Gemelli University Hospital, Università Cattolica del
46 Sacro Cuore, Roma 168, Italy.
- 47 160. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford
48 OX3 7FZ, UK.
- 49 161. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA.
- 50 162. Alzheimer's Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK.
- 51 163. Department of Clinical Sciences and Orthopaedic Surgery, Lund University, Malmö 20502, Sweden.
- 52 164. Imperial College Healthcare NHS Trust, London W12 0HS, UK.
- 53 165. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore.
- 54 166. MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK.
- 55 167. Unit of Primary Care, Oulu University Hospital, Oulu 90220, Finland.
- 56 168. Department of Pharmacology and Therapeutics, University College Cork, Cork T12 YT20, Ireland.

- 1 169. DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich 80336,
2 Germany.
- 3 170. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for
4 Environmental Health, Neuherberg 85764, Germany.
- 5 171. Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Neuherberg 85764, Germany.
- 6 172. Institute of Human Genetics, Technische Universität München, Munich 85764, Germany.
- 7 173. Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg 85764, Germany.
- 8 174. Munich Cluster for Systems Neurology (SyNergy), Munich 81377, Germany.
- 9 175. Information-based Medicine Stream, Hunter Medical Research Institute, New Lambton Heights, NSW
10 2305, Australia.
- 11 176. Harvard Medical School, Boston, MA 02115, USA.
- 12 177. Population Health Research Institute, St George's, University of London, London SW17 0RE, UK.
- 13 178. Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne 1011,
14 Switzerland.
- 15 179. Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of
16 Public Health, Imperial College London, London W2 1PG, UK
- 17 180. Durrer Center for Cardiogenetic Research, Amsterdam 3501 DG, the Netherlands.
- 18 181. Interuniversity Cardiology Institute of the Netherlands, Utrecht 3511 EP, the Netherlands.
- 19 182. Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald
20 17475, Germany.
- 21 183. Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New
22 Orleans, LA 70112, USA.
- 23 184. Analytical Genetics Group, Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY
24 10591, USA.
- 25 185. Institute for Translational Genomics and Population Sciences, Departments of Pediatrics and Medicine, Los
26 Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.
- 27 186. Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim 68161, Germany.
- 28 187. Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz 8036,
29 Austria.
- 30 188. Preventive Cardiology and Preventive Medicine, Center for Cardiology, University Medical Center of the
31 Johannes Gutenberg-University Mainz, Mainz 55131, Germany.
- 32 189. Department of Clinical Medicine, University of Copenhagen, Copenhagen 2300, Denmark.
- 33 190. Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
- 34 191. Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University
35 Hospital, Helsinki 00014, Finland.
- 36 192. MRC Human Genetics Unit, Institute for Genetics and Molecular Medicine, University of Edinburgh,
37 Western General Hospital, Edinburgh EH4 2XU, Scotland.
- 38 193. Departments of Neurology and Public Health Sciences, University of Virginia Charlottesville,
39 Charlottesville, VA 22908-0394, USA.
- 40 194. Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA
- 41 195. Renal Unit, Department of Medicine, Verona University Hospital, Verona 37126, Italy.
- 42 196. School of Public Health, University of Adelaide, Adelaide SA 5000, Australia.
- 43 197. Department of Pathology, University of California San Diego, San Diego, CA 92121, USA.
- 44 198. General Central Hospital, Department of Neurology, Bolzano 39100, Italy.
- 45 199. Department of Neurology, University of Lübeck, Lübeck 23538, Germany.
- 46 200. Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence 50122, Italy.
- 47 201. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala
48 University, Uppsala 751 41, Sweden.
- 49 202. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine,
50 Stanford, CA 94305, USA.
- 51 203. Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK.
- 52 204. Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI 48109,
53 USA.
- 54 205. National Heart and Lung Institute, Imperial College London, London W12 0NN, UK.
- 55 206. University of Kentucky, College of Public Health, Lexington, KY 40508, USA.

- 1 207. Department of Genomics of Common Disease, School of Public Health, Imperial College London, London
2 SW7 2AZ, UK.
3 208. Department of Clinical Epidemiology, Bispebjerg and Frederiksberg Hospital, Frederiksberg 2200,
4 Denmark.
5 209. Department of Internal Medicine II-Cardiology, University of Ulm Medical Center, Ulm 80636, Germany.
6 210. Deutsches Herzzentrum München, Technische Universität München, Munich 80636, Germany.
7 211. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York,
8 NY 10029-6542, USA.
9 212. Department of Neurology, Medical University Graz, Graz 8010, Austria.
10 213. Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK.
11 214. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA
12 98101, USA.
13 215. Department of Epidemiology, University of Washington, Seattle, WA 98101, USA.
14 216. Department of Health Services, University of Washington, Seattle, WA 98195-7660, USA.
15 217. Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA.
16 218. Department of Pathology, University of Vermont, Colchester, VT 05405, USA.
17 219. Human Genetics Center, School of Public Health, The University of Texas Health Science Center at
18 Houston, Houston, TX 77030, USA.
19 220. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.
20 221. Department of Lab Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA.
21 222. National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA 01702, USA.
22 223. Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA.

23

24 ***Corresponding authors**

25 Abbas Dehghan MD PhD

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

28 Room 157, Norfolk Place, St Mary's Campus

29 Phone: 0044 20 7594 3347

30 Email: a.dehghan@imperial.ac.uk

31

32 Behrooz Z. Alizadeh MD PhD

33 Department of Epidemiology, University Medical Center Groningen

34 Hanzeplein 1, 9713 GZ Groningen, the Netherlands

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

36 Email: b.z.alizadeh@umcg.nl

1 **Abstract**

2 C-reactive protein (CRP) is a sensitive biomarker of chronic low-grade inflammation that is associated
3 with multiple complex diseases. The genetic determinants of chronic inflammation remain largely
4 unknown, and the causal role of CRP in several clinical outcomes is debated. We performed two genome-
5 wide association studies (GWAS), HapMap and 1000Genomes imputed, of circulating CRP levels using
6 data from 88 studies comprising 204,402 European individuals. Additionally, we performed *in silico*
7 functional analyses and Mendelian randomization analyses with several clinical outcomes. The GWAS
8 meta-analyses of CRP revealed 58 distinct genetic loci ($P < 5 \times 10^{-8}$). After adjustment for body mass index
9 in the regression analysis, the associations at all except three loci remained. The lead variants at the
10 distinct loci explained up to 7.0% of the variance in circulating CRP levels. We identified 66 gene sets
11 that were organized in two substantially correlated clusters, one mainly comprised of immune pathways,
12 and the other characterized by metabolic pathways in the liver. Mendelian randomization analyses
13 revealed a causal protective effect of CRP on schizophrenia and a risk increasing effect on bipolar
14 disorder. Our findings provide further insights in the biology of inflammation that may lead to
15 interventions to treat inflammation and its clinical consequences.

1 **Introduction**

2 Inflammation plays a key role in the development of complex diseases such as cardiovascular disease¹,
3 type 2 diabetes², Alzheimer's disease³, and schizophrenia⁴. C-reactive protein (CRP) is a sensitive marker
4 of chronic low-grade inflammation⁵, and elevated serum levels of CRP have been associated with a wide
5 range of diseases⁶⁻⁸. Unraveling the genetics of inflammation may provide further insights into the
6 underlying biology of inflammation, and may identify therapeutic targets to attenuate inflammation.

7 The genetic determinants of CRP have only been partly characterized. In 2011, our group
8 published a HapMap-based meta-analysis of genome-wide association studies (GWAS) including a
9 discovery panel of up to 65,000 individuals and found 18 loci that were associated with CRP levels⁹.
10 Increasing study sample size in GWAS and denser mapping of the genome with further advanced
11 imputation panels may help to identify further genes associated with the phenotypes of interest^{10; 11}.
12 Furthermore, by using genetic instrumental variables (i.e., a genetic score), Mendelian randomization
13 (MR) allows investigation of the potential causal effect of an exposure on clinical outcomes, and may help
14 to understand the causal pathways that link the exposure with the outcome¹². The causal role of CRP in the
15 development of diseases is still controversial¹³, and the causal pathways that link inflammation to complex
16 disorders are only partly understood.

17 We applied two large-scale GWAS on circulatory levels of CRP using HapMap and
18 1000Genomes (1KG) imputed data to identify genetic determinants of chronic inflammation. Because
19 body mass index (BMI) is a major determinant of CRP levels, we additionally conducted GWAS adjusted
20 for BMI to identify associated loci independent of BMI. To identify any sex differences in genetic
21 determinants of chronic inflammation, we further conducted GWAS in men and women separately. We
22 applied *in silico* functional analyses on the identified loci to obtain better insights into the biological
23 processes potentially regulating chronic inflammation. Finally, we conducted MR analyses to provide an
24 improved understanding of the causal relation between CRP and several related clinical outcomes.

1 **Material and methods**

2 *GWAS for circulating CRP levels*

3 We conducted a meta-analysis of GWAS including individuals of European ancestry within the Cohorts
4 for Heart and Aging Research in Genomic Epidemiology consortium Inflammation Working Group of the
5 (CIWG)¹⁴. The CIWG invited cohorts for participation in the HapMap GWAS meta-analysis of CRP
6 levels in 2012. In 2014 and in light of our assessment which showed complementary value of HapMap
7 and 1KG imputed GWAS¹⁰, we invited studies to participate in the 1KG GWAS meta-analysis. The 1KG
8 GWAS may help to identify loci that were not covered in the HapMap GWAS and fine map loci found in
9 the HapMap GWAS. Cohorts were allowed to participate in either the HapMap or 1KG GWAS, or both.
10 Here we present both a HapMap (204,402 individuals from 78 studies) and 1KG (148,164 individuals
11 from 49 studies) imputed genotypes GWAS meta-analysis. All participating cohorts implemented a pre-
12 specified study plan comprising study design, data quality check, data analysis, and data sharing. We
13 measured serum CRP in mg/L using standard laboratory techniques (Supplemental Methods), and natural
14 log-transformed the values. Individuals with auto-immune diseases, individuals taking immune-
15 modulating agents (if this information was available), and individuals with CRP levels 4SD or more away
16 from the mean were excluded from all analyses. The characteristics of the participants are presented in
17 Table S1. We filtered individuals and genetic variants based on study-specific quality control criteria
18 (Table S2). At each individual study site, we tested genetic variants for association with CRP levels using
19 an additive linear regression model adjusted for age, sex, and population substructure, and accounting for
20 relatedness, if relevant. Before meta-analysis, we filtered variants based on imputation quality at R^2 index
21 of >0.4 . To avoid type-I error inflation, we corrected study-specific GWAS for genomic inflation. For the
22 HapMap study, we conducted fixed effect meta-analyses for each genetic variant, using the inverse
23 variance-weighted method implemented in GWAMA¹⁵, and performed a second genomic control on the
24 meta-analysis level. For the 1KG imputed GWAS, we used METAL¹⁶ to perform a fixed effect meta-
25 analysis. We removed variants that were only available in $<50\%$ of the samples. The HapMap meta-
26 analysis included 2,254,727 variants, and the 1KG GWAS included 10,019,203 variants. We considered

1 associations with $P < 5 \times 10^{-8}$ genome-wide significant. We used a stringent distance criterion, 500kb
2 minimum between two significant variants, to identify distinct loci. In each locus, the variant with the
3 smallest p-value was called the “lead variant”. Additionally, we performed sex-stratified analyses among
4 HapMap imputed studies, and we tested for heterogeneity between sex-specific effect estimates as
5 described previously¹⁷, using the false-discovery rate of Benjamini-Hochberg to assess significance of the
6 P for sex difference (< 0.05). We conducted BMI adjusted analyses in the IKG meta-analysis to determine
7 the role of BMI in mediating the genetic associations with CRP, and to increase power to detect
8 associations not mediated by BMI.

9

10 *LD Score regression*

11 Because population stratification is a major concern in GWAS and may lead to false-positive associations,
12 we applied Linkage Disequilibrium Score regression (LDSC) to distinguish whether the inflation of test
13 statistics observed in the CRP GWAS is due to the polygenic architecture of CRP or reflects confounding
14 bias due to cryptic relatedness or population stratification. The LD Score measures collective genetic
15 variation acquired from all genetic variants in LD with the index tagging (causal) variant¹⁸. A higher LD
16 score of an index variant implicates more nearby genetic variants in high LD with the index variant, which
17 makes it more likely that the index variant tags causal variant(s). More genetic variants in LD with the
18 index genetic variant (i.e., a higher LD score due to polygenicity) may yield higher (i.e., inflated) test
19 statistics. In contrast, higher test statistics caused by cryptic population stratification will not be related to
20 LD score. LD Score regression analysis performs regression of the summary statistics from the GWAS
21 meta-analysis (χ^2 statistics from the GWAS) on the LD scores across the genome. An intercept of the LD
22 Score regression that equals one suggests no confounding bias, whereas an inflated intercept (larger than
23 one) suggests contribution of confounding due to relatedness to the test statistics. We used the LDHub
24 web interface to perform LD Score regression¹⁹. We filtered variants to the subset of HapMap 3 variants,
25 and excluded variants with duplicated rs numbers, ambiguous variants, minor allele frequency

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

3 Furthermore, we applied cross-trait LD score regression to estimate genetic correlation of chronic
4 inflammation (using the HapMap GWAS meta-analysis) with other phenotypes using published GWAS
5 summary statistics²⁰. In brief, the cross-product of two GWAS test statistics is calculated at each genetic
6 variant, and this cross-product is regressed on the LD Score. The slope of the regression is used to
7 estimate the genetic covariance between two phenotypes.

8

9 *Identification of additional distinct variants in associated loci*

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

20

21 *Proportion of CRP variance explained*

22 We estimated the variance explained in serum CRP levels using the formula $(2 * MAF(I -$
23 $MAF)beta^2) / var(CRP)$, where beta is the estimated effect of the individual variants on CRP²³ and
24 $var(CRP)$ is the variance in natural log-transformed CRP estimated in the RS-I cohort. We calculated the
25 variance explained for four combinations of associated variants: 1. the lead variant at just the *CRP* locus;
26 2. the distinct variants at the CRP locus derived from the 1KG joint conditional analysis; 3. all lead

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

3

4 *Pathway analysis and gene expression*

5 We used Data-Driven Expression-Prioritized Integration for Complex Traits²⁴ (DEPICT v.1 rel173 beta)
6 to systematically prioritize the most likely causal genes, highlight the pathways that are enriched by the
7 likely causal genes and identify tissues and cell types in which genes from associated loci are highly
8 expressed. DEPICT requires summary statistics from the GWAS meta-analysis. First, we filtered genome-
9 wide associated variants from both GWAS meta-analyses by $MAF > 0.01$, and selected variants with low
10 correlation with other variants by PLINK (version 1.90) using a clumping distance of 500 kb apart and/or
11 index of LD r^2 threshold < 0.1 . The settings for the analysis involved the usage of 1KG pilot phase data²⁵
12 (phase 1 integrated release, version 3, CEU, GBR, TSI unrelated individuals; 2010.11.23) with $r^2 > 0.5$ LD
13 threshold for locus definition, 10,000 permutations for bias correction, and 500 repetitions for FDR
14 calculation. To summarize and visualize the results, we calculated pairwise Pearson correlation
15 coefficients between all gene-specific Z-scores for every pair of reconstituted DEPICT gene sets. We used
16 Affinity Propagation Clustering (apcluster command; *APCluster* R package²⁶) to identify clusters and
17 representative examples of the clusters, and Cytoscape v3.2.1 for visualization of the results. The DEPICT
18 results of the pathway and gene prioritization are summarized as a heatmap (R. v2.3.3, *heatmap* v1.0.8
19 package²⁷). The gene-specific Z-score describes the likelihood that a given gene is part of the
20 corresponding GO term, KEGG pathway, REACTOME pathway, Mouse Phenotype, or protein-protein
21 interaction network.

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

1 To prioritize the most likely trait-relevant gene for each GWAS locus, we run colocalization
2 analysis using the “coloc” R package v3.1³⁰ separately for the HapMap and 1KG GWAS. We used
3 publicly available genome-wide eQTL data from 5,311 whole blood samples³¹, and from the Genome
4 Tissue Expression (GTEx) V6p portal incorporating eQTL data from 44 post-mortem tissues³². “Coloc”
5 uses approximate Bayes factors to estimate the posterior probability that GWAS and eQTL effects share a
6 single causal variant. All significant cis-eGenes or cis-eProbes ($q < 0.05$ in GTEx; lowest cis-eQTL
7 FDR < 0.05 in Westra et al.³¹) were extracted ± 1 Mb from the lead SNP of each locus. The HapMap SNP
8 positions were converted to hg19 with the liftOver command from the rtracklayer v1.38.3 package. We
9 used the SNPs present in both the GWAS and eQTL datasets. For the HapMap GWAS, the 1KG GWAS
10 and the GTEx eQTL datasets, we performed the test using association beta, standard error of beta, and
11 minor allele frequency (MAF). For the data from Westra et al.³¹, we used association P-value, MAF, and
12 sample size, and included only the subset of cis-eQTLs which are publicly available (up to significance
13 FDR < 0.5). We used default priors supplied by the coloc package ($P_1 = 1e-4$, $P_2 = 1e-4$, $P_{12} = 1e-5$; prior
14 probabilities for association in GWAS, eQTL, and both datasets). Full MAF data were not available for
15 the eQTL datasets, therefore we used the GIANT 1KG p1v3 EUR reference panel instead. We visualized
16 the results as a heatmap using the *pheatmap* v1.0.8 R package.²⁷

17

18 *Mendelian randomization*

19 To assess the effect of CRP on complex disorders, we performed a two-sample Mendelian randomization
20 (MR) study on nine clinical outcomes (Alzheimer’s disease (AD), bipolar disorder (BD), coronary artery
21 disease (CAD), Crohn’s disease (CD), inflammatory bowel disease (IBD), rheumatoid arthritis (RA),
22 schizophrenia, and diastolic (DBP) and systolic blood pressure (SBP)) to which CRP showed a potentially
23 causal association at a $P < 0.1$ in a previous MR study¹³. We used the effect estimates of the 48 lead SNPs
24 found to be associated with CRP in the HapMap GWAS, and the effect estimates of the four SNPs that
25 were additionally found to be associated with CRP in the 1KG GWAS in a multiple instrument approach
26 for the MR analyses ($n = 52$ SNPs). Additionally, we separately studied the effect of rs2794520 at the *CRP*

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

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

4
5 R^2 is the variance explained of CRP by the genetic instrument (0.014 for the rs2794520 SNP and 0.065 for
6 the 52-SNP score), n is the number of individuals included in the CRP GWAS, and k the number of
7 variants included in the genetic score. The F statistic for the 52-SNP score was 273, and for the rs2794520
8 SNP 2,902, indicating that both instruments were strong.

9 For the clinical outcomes, we used summary statistics from the most recent meta-analysis of
10 GWA studies. For diastolic and systolic blood pressure, we used data from the UK Biobank. The details of
11 the outcome studies are summarized in Table S12. We implemented four different methods of MR
12 analyses: Inverse-variance weighted method (IVW), MR-Egger, Weighted median (WM), and Penalized
13 weighted median (PWM). We used the “*TwoSampleMR*” package in R for the MR analyses³³. Further, we
14 applied the Bonferroni method to correct for multiple testing ($0.05/9$ phenotypes = 5.6×10^{-3}). When the Q-
15 statistic of the IVW analyses provided evidence for heterogeneity, the weighted median estimates were
16 used for significance. The MR methods are described briefly below.

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

24 MR-Egger: We used MR-Egger to account for potential unbalanced pleiotropy in the multiple
25 variant instrument³⁴. When unbalanced pleiotropy is present, an alternative effect (positive or negative) is

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

9 Weighted median (WM) and penalized Weighted Median (PWM): We applied the median based
10 method to provide robust estimates of causal association even in the presence of horizontal pleiotropy
11 when up to 50% of the information contributed by the genetic variants is invalid³⁵. In PWM analysis the
12 effect of each variants is weighted by a factor that corresponds to the Q statistics (heterogeneity test) of
13 the SNP; this means that most variants will not be affected by this correction, but the causal effect of the
14 outlying variants, which are most likely to be invalid IVs, will be down-weighted.

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

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

1

2 **Results**

3 *HapMap GWAS meta-analysis for CRP levels*

4 The HapMap meta-analysis identified 3,977 genome-wide significant variants at $P < 5 \times 10^{-8}$ (QQ-plot
5 Figure S1; Manhattan plot Figure S2), which mapped to 48 distinct loci (Table 1, Table S3). Of the
6 previously reported 18 variants for CRP, 16 remained associated. Compared to the previous GWAS, the
7 rs6901250 variant at the *GPRC6A* locus ($P=0.09$) and the rs4705952 variants at the *IRF1* locus ($P=2.7 \times 10^{-3}$)
8 were not significant. The beta estimates for natural log-transformed CRP for each of the associated loci
9 ranged from 0.020 to 0.229. We observed the strongest association for rs2794520 at the *CRP* gene
10 ($\beta=0.182$ in the natural log-transformed CRP (mg/L) per copy increment in the coded allele, $P=4.17 \times 10^{-523}$),
11 followed by rs4420638 at the *APOC1/E* gene ($\beta=0.229$, $P=1.23 \times 10^{-305}$). Similarly to previous GWAS
12 meta-analysis, the lead variant within the interleukin-6 receptor gene (*IL6R*) was rs4129267 ($\beta=0.088$,
13 $P=1.2 \times 10^{-129}$). Related to the interleukin-6 pathway, we identified rs1880241 upstream of the *IL6* gene ($\beta=$
14 0.028 , $P=8.4 \times 10^{-14}$). In addition to the previously described interleukin-1 signaling, the *IL1RN-IL1F10*
15 locus (interleukin-1 receptor antagonist and interleukin-1 family member 10), we found rs9284725 within
16 the interleukin-1 receptor 1 gene (*IL1RI*, $\beta=0.02$, $P=7.3 \times 10^{-11}$, Table 1). The sex-specific meta-analyses
17 did not identify additional loci for CRP compared to the overall meta-analysis including both sexes, but at
18 four genetic variants we found evidence for heterogeneity in effect estimates between sexes (Table S4),
19 though the directions of associations were consistent.

20

21 *1KG GWAS meta-analysis for CRP levels*

22 In the 1KG meta-analysis, 8,002 variants were associated with CRP at $P < 5 \times 10^{-8}$ (QQ-plot Figure S3;
23 Manhattan plot Figure S4). This resulted in 40 distinct loci, of which 36 overlapped with the HapMap
24 meta-analysis (Table 1). The lead variant at the *CRP* locus in the 1KG GWAS was rs4287174 ($\beta=-0.185$,
25 $P=1.95 \times 10^{-398}$), which is in high LD with rs2794520 ($r^2=0.98$), the lead variant at the *CRP* locus in the
26 HapMap GWAS. Among eight of the overlapping loci, the lead variant was at the same position in both

1 GWAS (rs1260326, rs1490384, rs10832027, rs1582763, rs7310409, rs2239222, rs340005, and
2 rs1800961). Compared with HapMap, the four additional variants identified in 1KG were rs75460349,
3 rs1514895, rs112635299, and rs1189402. The variants rs1514895 and rs1189402 were available in the
4 HapMap GWAS, but were not associated at the genome-wide threshold (respectively $P=1.2\times 10^{-7}$ and
5 $P=8.1\times 10^{-3}$). The two variants rs75460349 and rs112635299 were not available in the HapMap GWAS,
6 nor were variants in high LD ($r^2<0.8$). The rs75460349 is a low frequency variant with a coded allele
7 frequency of 0.97 ($\beta=0.086$, $P=4.5\times 10^{-10}$). Also rs112635299 near the *SERPINA1/2* gene is a low
8 frequency variant with a MAF of 0.02 ($\beta=0.107$, $P=2.1\times 10^{-10}$). Adjustment for BMI in the 1KG GWAS
9 ($n=147,827$) revealed six additional loci that were not associated with CRP in the HapMap and 1KG
10 primary analyses (Table 1; Table S5, QQ-plot Figure S5; Manhattan plot Figure S6). The associations at
11 three lead variants were much reduced after adjustment for BMI (rs1558902 (*FTO*), $P_{adjusted}=0.40$;
12 rs12995480 (*TMEM18*), $P_{adjusted}=0.02$; rs64343 (*ABO*), $P_{adjusted}=1.0\times 10^{-7}$). Both the *FTO* and *TMEM18*
13 gene are well-known obesity genes. Except for the *FTO*, *TMEM18*, and *ABO* loci, all distinct loci
14 identified in the primary 1KG analysis were also associated with CRP in the BMI adjusted 1KG analysis.
15 No genome-wide significant association was observed on the X-chromosome in the 1KG GWAS
16 including 102,086 individuals.

17

18 *LD score regression*

19 The HapMap GWAS LD Score regression intercept was 1.03 (standard error: 0.013), and the 1KG
20 intercept was 1.02 (standard error 0.011). This suggests that a small proportion of the inflation is
21 attributable to confounding bias (~12% for the HapMap GWAS and ~13% for the 1KG GWAS). Hence,
22 the vast majority of inflation is due to the polygenic architecture of circulating CRP levels. As depicted in
23 Figure 1, CRP showed strong positive genetic correlations with anthropometric traits (e.g., BMI: $R_g=0.43$,
24 $P=5.4\times 10^{-15}$), glycemic phenotypes (e.g., type 2 diabetes $R_g=0.33$, $P=3.1\times 10^{-6}$), lipid phenotypes (e.g.,
25 triglycerides $R_g=0.29$, $P=7.9\times 10^{-5}$), and coronary artery disease ($R_g=0.23$, $P=2.4\times 10^{-5}$) (Table S6). By
26 comparison, CRP showed inverse genetic correlations with educational attainment (e.g., college

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

3
4 *Additional signals at distinct loci*
5 Approximate conditional analyses in the 1KG GWAS revealed additional signals at nine loci (Table S7).
6 Five loci showed one secondary signal (*IL6R*, *NLRP3*, *HNF1A*, *CD300LF*, and *APOE/APOC1*), the
7 *PPP1R3B* locus had two additional signals, the *LEPR* locus had three additional signals, and the *SALL1*
8 locus had four additional signals, whereas the *CRP* locus showed a total of 13 distinct associated variants.
9 Interestingly, the rs149520992 rare variant (MAF=0.01) mapping to the *CRP* locus showed an association
10 at $P_{\text{conditional}}=3.7\times 10^{-15}$ with $\beta=-0.272$ for the T-allele. The GCTA effect estimates for the ten distinct
11 variants in the vicinity of the *CRP* gene identified in the 1KG conditional analysis are in high correlation
12 with the effect estimates of these variants obtained from the RS-I and WGHS individual level data
13 ($r_{\text{RS}}=0.97$, and $r_{\text{WGHS}}=0.84$), confirming the reliability of the GCTA estimates.

14
15 *Variance explained of CRP*

16 The lead variant at the *CRP* locus in both the HapMap (rs2794520) and 1KG (rs4287174) GWAS
17 explained 1.4% of the variance in natural log-transformed CRP levels. The distinct variants at the *CRP*
18 locus derived from the joint conditional analysis in the 1KG GWAS explained 4.3% of the variance. The
19 lead variants at all distinct loci together explained 6.2% of the CRP variance in the HapMap GWAS, and
20 6.5% in the 1KG GWAS. When we added the distinct variants at associated loci derived from the
21 conditional analysis, the variance explained by all associated loci was 11.0% in the 1KG GWAS.

22
23 *Functional annotation*

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

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

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

23 24 *Mendelian randomization analyses*

25 We observed a protective effect of genetically determined variance in CRP with schizophrenia with an
26 IVW odds ratio (OR) of the 52-SNP score of 0.89 (95%CI: 0.81-0.97, P=6.6×10⁻³) (Tables S14-S15,

1 Figure S8-S11). The MR-Egger intercept was compatible with no unbalanced pleiotropy ($P=0.48$). The
2 estimate of the rs2794520 variant was comparable to the 52-SNP score estimate (OR 0.89, 95% CI 0.84-
3 0.94, $P=0.046$). The WM and PWM estimates were comparable to the IVW estimate (OR_{WM} 0.89,
4 $P_{WM}=5.1\times 10^{-3}$; OR_{PWM} 0.89, $P_{PWM}=4.4\times 10^{-3}$). The “leave-one-out” analysis provided evidence that no
5 single variant was driving the IVW point estimate (Figure S10). The causal OR between the rs2794520
6 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
7 (95% CI 1.00-1.35, $P=0.054$). The MR-Egger intercept was compatible with unbalanced pleiotropy
8 ($P=0.049$). The MR-Egger estimate OR of the 52-SNP score was comparable to the rs2794520 estimate
9 (OR 1.36, 95%CI 1.1.-1.69, $P=6.7\times 10^{-3}$), as were the WM and PWM estimates (OR_{WM} 1.33, $P_{WM}=3.4\times 10^{-3}$;
10 OR_{PWM} 1.32, $P_{PWM}=4.3\times 10^{-3}$).

11 We observed evidence against a causal association between either *CRP* rs2794520 (OR 1.01,
12 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
13 Egger intercept of 0.014 suggested presence of unbalanced pleiotropy ($P=5.8\times 10^{-3}$), with an MR-Egger
14 causal estimate of OR 0.79 (95%CI 0.67-0.94, $P=0.012$). However, the WM and PWM showed no
15 association between CRP and CAD. For AD, there was evidence against an association with rs2794520
16 ($P=0.592$), though the IVW OR showed a protective effect (OR 0.51, 95%CI 0.30-0.88, $P=0.015$). The
17 Egger intercept of 0.046 suggested unbalanced pleiotropy ($P=0.042$), and the MR-Egger OR was 0.27
18 (95%CI 0.12-0.60). However, the association was null for the WM and PWM analyses (OR_{WM} 1.04,
19 $P_{WM}=0.61$; OR_{PWM} 1.05, $P_{PWM}=0.53$). We observed evidence against an effect for CD, DBP, IBD, RA, and
20 SBP for the rs2794520 variant and the IVW, MR-Egger, WM, and PWM analyses.

21

22 Discussion

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

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

5 Obesity is one of the main determinants of chronic low-grade inflammation in the general
6 population^{37, 38}. Adjustment for BMI in the CRP GWAS abolished the association at only three lead
7 variants, suggesting that the genetic regulation of chronic low-grade inflammation is largely independent
8 from BMI. Notably, BMI adjustment resulted in the identification of six variants that were not associated
9 with CRP in the BMI-unadjusted GWAS. This supports the notion that adjustment for covariates that
10 explain phenotypic variance may improve the statistical power in linear model analyses of quantitative
11 traits³⁹. Although adjustment for heritable correlated traits in GWAS may bias effect estimates (collider
12 bias)⁴⁰, there is consistent evidence in the literature that BMI has a causal direct effect on CRP levels⁴¹,
13 and therefore, collider bias in CRP GWAS adjusted for BMI is less likely.

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

21 The top variant at the *CRP* locus in both the HapMap and 1KG GWAS explained 1.4% of the
22 variance in circulating CRP levels. The approximate conditional analysis resulted in 13 variants jointly
23 associated within the *CRP* locus in the 1KG GWAS. With respect to locus definition, we used a
24 conservative distance criterion compared to other GWA studies that often use $\pm 500\text{kb}$ surrounding the
25 GWAS peak⁴². Here, we used the criterion that the minimum distance between the boundaries of loci is
26 500kb. In order to identify further variants associated with CRP levels, we performed approximate

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

9 DEPICT analysis provided further evidence that the genes annotated to the associated CRP
10 variants mainly cluster in the immune and liver biological systems. Notably, the gene set “inflammatory
11 response”, which captures both immune response and liver metabolism, was the main connector network
12 between the two networks. This is in line with the observation that CRP is mainly produced by liver cells
13 in response to inflammatory cytokines during acute and chronic inflammation⁴³. Interestingly, the analysis
14 highlighted iron homeostasis as an enriched gene set. In agreement, the conditional analysis highlighted a
15 distinct genetic association at the hemochromatosis gene *HFE*, a transmembrane protein of the major
16 histocompatibility complex (MHC) class I family. Previous studies show that iron metabolism plays a
17 pivotal role in inflammation^{44; 45}. However, genetic pleiotropy may highlight co-regulated networks in
18 pathway analysis that are not causal to inflammation per se. It is also important to note that the results of
19 DEPICT analyses apply to reconstituted gene sets which may sometimes have slightly different overlaying
20 biological theme than the original gene set annotation.

21 The MR analyses validate previous evidence that genetically-elevated CRP is protective for the
22 risk of schizophrenia^{13; 46}, although observational data suggest a positive association between CRP and
23 risk of schizophrenia⁴⁷. For bipolar disorder we observed a positive causal effect, which is in line with
24 previous MR and observational studies^{13; 48}. Although the causal underlying mechanisms remain to be
25 elucidated, a hypothesis for the schizophrenia observation might be the immune response to infections
26 early in life. Levels of acute-phase response proteins in dry blood spots collected at birth are lower for

1 patients with non-affective psychosis, which includes schizophrenia, compared to controls, suggesting a
2 weaker immune response at birth⁴⁹. Also, neonates that have been exposed to a maternal infection and
3 have low levels of acute-phase response proteins, have a higher risk of schizophrenia⁵⁰. Altogether, the
4 evidence suggests that a deficient immune response may contribute to chronic infection in children and the
5 development of schizophrenia. For AD and CAD, the Egger intercept showed evidence of unbalanced
6 pleiotropy and the Egger estimate showed a protective effect of CRP on the risk of AD and CAD.
7 However, for both outcomes, the effects of the WM and PWM analyses, as well as analyses using the
8 single rs2794520 variant (which is least likely to be affected by pleiotropy) were null. The MR-Egger
9 estimate relies on the InSIDE assumption which states that the strength of the association between the
10 genetic variants and CRP is independent from the strength of the direct pleiotropic effects of the genetic
11 variants on the outcome. This assumption may be violated when the genetic variants are associated with a
12 confounder of the CRP-outcome association. Such a scenario may occur when the genetic variants are
13 associated with an exposure that is causally upstream of the exposure under study. In the context of the
14 association of CRP with AD and CAD, this could be lipids or glycemic phenotypes. Several genetic
15 variants used in the 52-SNP instrument are associated with metabolic phenotypes that may affect CRP
16 levels. In agreement, the WM and PWM, in which the InSIDE assumption is relaxed, and the single
17 variant analysis showed no association. Furthermore, the observation that CRP is not causally related to
18 CAD in the MR analyses is in comparison to previous published studies⁵¹. Power calculation showed that
19 we had 100% power to detect a 10% difference in CAD risk, thus the probability of a false negative
20 finding is small. Also, CRP is associated with future CAD in observational studies, and randomized trials
21 have shown a beneficial effect of lowering inflammation using statins⁵² and canakinumab⁵³ on CAD risk,
22 but this effect is unlikely to be attributable to CRP.

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

1 without replication. The consistent association of the variants in >50 studies at a strict Bonferroni
2 corrected threshold provide confidence that our findings represent true associations. We used both
3 HapMap and 1KG imputed data to identify genetic variants for circulating CRP levels. At the start of the
4 project, more studies had HapMap imputed data available. Hence, the sample size and thus power in the
5 HapMap GWAS was higher compared to the 1KG. Also, HapMap may identify variants that are not
6 identified in 1KG GWAS⁵⁴. Nevertheless, 1KG offers better coverage of uncommon variants and includes
7 INDELS, which are not included in the HapMap reference panel. Including both reference panels, we used
8 all available samples and maximized the possibility to identify genetic variants for CRP, both common
9 and uncommon.

10 However, we note limitations to our study. GWAS merely identify loci associated with complex
11 phenotypes and the identification of causal genes remains challenging. We only included individuals of
12 European ancestry; the generalizability of our findings to other races/ethnicities is uncertain. In addition,
13 although our analyses provided support for causal associations, we acknowledge that we may not have
14 identified the causal variants and we may not have eliminated residual confounding. The colocalisation
15 analyses provide evidence for colocalisation of CRP GWAS signals and eQTLs, however, it does not
16 provide evidence that the GWAS signal is functioning on CRP through the gene expression. We further
17 note that the method assumes identical LD-structure from the GWAS and eQTL datasets. As there are
18 ~14% of non-European samples in the full GTEx dataset, this assumption might be violated for some
19 tissues. Last, we meta-analyzed all available samples in one meta-analysis and did not replicate our
20 findings in an independent sample. Therefore, our findings may need replication.

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

26

1 **Supplemental data**

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

3

4 **Acknowledgements**

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

6

7 **Declaration of interest**

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

12

13 **Web Resources**

14 DEPICT, <https://data.broadinstitute.org/mpg/depict/>

15 GCTA, cnsgenomics.com/software/gcta/

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

18 GTEx Portal, <https://www.gtexportal.org/>

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

20 LD Hub, ldsc.broadinstitute.org/ldhub/

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

22 [mRnd power calculations for Mendelian Randomization, http://cnsgenomics.com/shiny/mRnd/](https://cnsgenomics.com/shiny/mRnd/)

23

24 **Summary data**

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

1

2 References

- 3 1. Libby, P. (2012). Inflammation in atherosclerosis. *Arterioscler Thromb Vasc Biol* 32, 2045-2051.
- 4 2. Pickup, J.C. (2004). Inflammation and activated innate immunity in the pathogenesis of type 2
5 diabetes. *Diabetes Care* 27, 813-823.
- 6 3. Akiyama, H., Barger, S., Barnum, S., Bradt, B., Bauer, J., Cole, G.M., Cooper, N.R., Eikelenboom,
7 P., Emmerling, M., Fiebich, B.L., et al. (2000). Inflammation and Alzheimer's disease. *Neurobiol*
8 *Aging* 21, 383-421.
- 9 4. Khandaker, G.M., Cousins, L., Deakin, J., Lennox, B.R., Yolken, R., and Jones, P.B. (2015).
10 Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment.
11 *Lancet Psychiatry* 2, 258-270.
- 12 5. Pepys, M.B. (1995). The acute phase response and C-reactive protein. *Oxford textbook of medicine*
13 2, 1527-1533.
- 14 6. Danesh, J., Wheeler, J.G., Hirschfield, G.M., Eda, S., Eiriksdottir, G., Rumley, A., Lowe, G.D.,
15 Pepys, M.B., and Gudnason, V. (2004). C-reactive protein and other circulating markers of
16 inflammation in the prediction of coronary heart disease. *N Engl J Med* 350, 1387-1397.
- 17 7. Dehghan, A., Kardys, I., de Maat, M.P., Uitterlinden, A.G., Sijbrands, E.J., Bootsma, A.H., Stijnen,
18 T., Hofman, A., Schram, M.T., and Witteman, J.C. (2007). Genetic variation, C-reactive protein
19 levels, and incidence of diabetes. *Diabetes* 56, 872-878.
- 20 8. Wium-Andersen, M.K., Ørsted, D.D., and Nordestgaard, B.G. (2013). Elevated C-reactive protein
21 associated with late-and very-late-onset schizophrenia in the general population: a prospective study.
22 *Schizophr Bull*, sbt120.
- 23 9. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski,
24 H., Kettunen, J., Henneman, P., et al. (2011). Meta-analysis of genome-wide association studies in
25 >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* 123, 731-738.
- 26 10. de Vries, P.S., Sabater-Lleal, M., Chasman, D.I., Trompet, S., Ahluwalia, T.S., Teumer, A., Kleber,
27 M.E., Chen, M.H., Wang, J.J., Attia, J.R., et al. (2017). Comparison of HapMap and 1000 Genomes
28 Reference Panels in a Large-Scale Genome-Wide Association Study. *PLoS One* 12, e0167742.
- 29 11. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017).
30 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 101, 5-22.
- 31 12. Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N., and Davey Smith, G. (2008). Mendelian
32 randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*
33 27, 1133-1163.
- 34 13. Prins, B.P., Abbasi, A., Wong, A., Vaez, A., Nolte, I., Franceschini, N., Stuart, P.E., Gutierrez
35 Achury, J., Mistry, V., Bradfield, J.P., et al. (2016). Investigating the causal relationship of C-reactive
36 protein with 32 complex somatic and psychiatric outcomes: a large-scale cross-consortium Mendelian
37 randomization study. *PLoS Med* 13, e1001976.
- 38 14. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden,
39 A.G., Harris, T.B., Witteman, J.C.M., and Boerwinkle, E. (2009). Cohorts for heart and aging
40 research in genomic epidemiology (CHARGE) consortium design of prospective meta-analyses of
41 genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2, 73-80.
- 42 15. Mägi, R., and Morris, A.P. (2010). GWAMA: software for genome-wide association meta-analysis.
43 *BMC Bioinformatics* 11, 288.
- 44 16. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of
45 genomewide association scans. *Bioinformatics* 26, 2190-2191.
- 46 17. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpelainen,
47 T.O., Esko, T., Magi, R., Li, S., et al. (2013). Sex-stratified genome-wide association studies
48 including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits.
49 *PLoS Genet* 9, e1003500.

- 1 18. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price,
2 A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in
3 genome-wide association studies. *Nat Genet* 47, 291-295.
- 4 19. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G.,
5 Tansey, K., Laurin, C., Pourcain, B.S., et al. (2017). LD Hub: a centralized database and web
6 interface to perform LD score regression that maximizes the potential of summary level GWAS data
7 for SNP heritability and genetic correlation analysis. *Bioinformatics* 33, 272-279.
- 8 20. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry,
9 J.R., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human
10 diseases and traits. *Nat Genet* 47, 1236-1241.
- 11 21. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide
12 complex trait analysis. *Am J Hum Genet* 88, 76-82.
- 13 22. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A., Heath, A.C., Martin, N.G.,
14 Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP
15 analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat*
16 *Genet* 44, 369-375.
- 17 23. Park, J.-H., Wacholder, S., Gail, M.H., Peters, U., Jacobs, K.B., Chanock, S.J., and Chatterjee, N.
18 (2010). Estimation of effect size distribution from genome-wide association studies and implications
19 for future discoveries. *Nat Genet* 42, 570-575.
- 20 24. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S.,
21 Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies
22 using predicted gene functions. *Nat Commun* 6, 5890.
- 23 25. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A.,
24 Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated
25 map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65.
- 26 26. Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity
27 propagation clustering. *Bioinformatics* 27, 2463-2464.
- 28 27. Kolde, R. (2012). Pheatmap: pretty heatmaps. R package version 61.
- 29 28. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set
30 analysis of GWAS data. *PLoS Comput Biol* 11, e1004219.
- 31 29. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and
32 annotation of genetic associations with FUMA. *Nat commun* 8, 1826.
- 33 30. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and
34 Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using
35 summary statistics. *PLoS Genet* 10, e1004383.
- 36 31. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen,
37 M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans
38 eQTLs as putative drivers of known disease associations. *Nat Genet* 45, 1238-1243.
- 39 32. GTEx Consortium. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene
40 regulation in humans. *Science* 348, 648-660.
- 41 33. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S.,
42 Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference
43 across the human phenome. *eLife* 7, e34408.
- 44 34. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid
45 instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 44, 512-
46 525.
- 47 35. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent estimation in
48 Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet*
49 *Epidemiol* 40, 304-314.
- 50 36. Brion, M.-J.A., Shakhbazov, K., and Visscher, P.M. (2012). Calculating statistical power in
51 Mendelian randomization studies. *Int J Epidemiol* 42, 1497-1501.

- 1 37. Visser, M., Bouter, L.M., McQuillan, G.M., Wener, M.H., and Harris, T.B. (1999). Elevated C-
2 reactive protein levels in overweight and obese adults. *JAMA* 282, 2131-2135.
- 3 38. Wellen, K.E., and Hotamisligil, G.S. (2003). Obesity-induced inflammatory changes in adipose
4 tissue. *J Clin Invest* 112, 1785.
- 5 39. Robinson, L.D., and Jewell, N.P. (1991). Some surprising results about covariate adjustment in
6 logistic regression models. *International Statistical Review/Revue Internationale de Statistique*, 227-
7 240.
- 8 40. Aschard, H., Vilhjálmsson, B.J., Joshi, A.D., Price, A.L., and Kraft, P. (2015). Adjusting for heritable
9 covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet* 96, 329-
10 339.
- 11 41. Timpson, N.J., Nordestgaard, B.G., Harbord, R.M., Zacho, J., Frayling, T.M., Tybjaerg-Hansen, A.,
12 and Smith, G.D. (2011). C-reactive protein levels and body mass index: elucidating direction of
13 causation through reciprocal Mendelian randomization. *Int J Obes* 35, 300-308.
- 14 42. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S.,
15 Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for
16 obesity biology. *Nature* 518, 197-206.
- 17 43. Moshage, H.J., Roelofs, H.M.J., Van Pelt, J.F., Hazenberg, B.P.C., Van Leeuwen, M.A., Limburg,
18 P.C., Aarden, L.A., and Yap, S.H. (1988). The effect of interleukin-1, interleukin-6 and its
19 interrelationship on the synthesis of serum amyloid A and C-reactive protein in primary cultures of
20 adult human hepatocytes. *Biochem Biophys Res Commun* 155, 112-117.
- 21 44. Ganz, T., and Nemeth, E. (2015). Iron homeostasis in host defence and inflammation. *Nat Rev*
22 *Immunol* 15, 500-510.
- 23 45. Alizadeh, B., Njajou, O., Hazes, J., Hofman, A., Slagboom, P., Pols, H., and van Duijn, C. (2007).
24 The H63D variant in the HFE gene predisposes to arthralgia, chondrocalcinosis and osteoarthritis.
25 *Ann Rheum Dis* 66, 1436-1442.
- 26 46. Hartwig, F.P., Borges, M.C., Horta, B.L., Bowden, J., and Smith, G.D. (2017). Inflammatory
27 Biomarkers and Risk of Schizophrenia: A 2-Sample Mendelian Randomization Study. *JAMA*
28 *psychiatry*.
- 29 47. Fernandes, B.S., Steiner, J., Bernstein, H.-G., Dodd, S., Pasco, J., Dean, O., Nardin, P., Goncalves,
30 C.-A., and Berk, M. (2016). C-reactive protein is increased in schizophrenia but is not altered by
31 antipsychotics: meta-analysis and implications. *Mol Psychiatry* 21, 554-565.
- 32 48. Fernandes, B.S., Steiner, J., Molendijk, M.L., Dodd, S., Nardin, P., Gonçalves, C.-A., Jacka, F.,
33 Köhler, C.A., Karmakar, C., Carvalho, A.F., et al. (2016). C-reactive protein concentrations across
34 the mood spectrum in bipolar disorder: a systematic review and meta-analysis. *Lancet Psychiatry* 3,
35 1147-1156.
- 36 49. Gardner, R., Dalman, C., Wicks, S., Lee, B., and Karlsson, H. (2013). Neonatal levels of acute phase
37 proteins and later risk of non-affective psychosis. *Transl psychiatry* 3, e228.
- 38 50. Blomström, Å., Gardner, R., Dalman, C., Yolken, R., and Karlsson, H. (2015). Influence of maternal
39 infections on neonatal acute phase proteins and their interaction in the development of non-affective
40 psychosis. *Transl psychiatry* 5, e502.
- 41 51. Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund,
42 P., Engert, J.C., and Bennett, D. (2009). Genetic loci associated with C-reactive protein levels and
43 risk of coronary heart disease. *JAMA* 302, 37-48.
- 44 52. Ridker, P.M., Danielson, E., Fonseca, F.A., Genest, J., Gotto, A.M., Jr., Kastelein, J.J., Koenig, W.,
45 Libby, P., Lorenzatti, A.J., MacFadyen, J.G., et al. (2008). Rosuvastatin to prevent vascular events in
46 men and women with elevated C-reactive protein. *N Engl J Med* 359, 2195-2207.
- 47 53. Ridker, P.M., Everett, B.M., Thuren, T., MacFadyen, J.G., Chang, W.H., Ballantyne, C., Fonseca, F.,
48 Nicolau, J., Koenig, W., Anker, S.D., et al. (2017). Antiinflammatory Therapy with Canakinumab for
49 Atherosclerotic Disease. *N Engl J Med* 377, 1119-1131.
- 50 54. Wood, A.R., Perry, J.R., Tanaka, T., Hernandez, D.G., Zheng, H.-F., Melzer, D., Gibbs, J.R., Nalls,
51 M.A., Weedon, M.N., Spector, T.D., et al. (2013). Imputation of variants from the 1000 Genomes

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

3

4 **Figure Legends**

5 **Figure 1. Genome-wide genetic correlation between serum CRP levels and different phenotypes and**

6 **clinical diseases.** The genetic correlation and its standard error are estimated with linkage disequilibrium

7 score regression analysis. ADHD, attention deficit and hyperactivity disorder; FEV1, forced expiratory

8 volume in 1 second; FVC, forced vital capacity; HOMA-B, homeostatic model assessment β -cell function;

9 HOMA-IR, homeostatic model assessment insulin resistance; HbA1C, Hemoglobin A1c.

10

11 **Figure 2. Results of the DEPICT functional annotation analysis.** Each node represents exemplar gene

12 set from Affinity Propagation clustering and links represent corresponding Pearson correlation coefficients

13 between individual enriched gene sets (only the links with $r > 0.3$ are shown). As an example, outlined are

14 the individual gene sets inside two clusters (“Inflammatory response” and “negative regulation of
15 peptidase activity”).

16

17 **Figure 3. Heatmap representing the results of DEPICT functional annotation analysis.** Each row

18 represents enriched ($FDR < 0.05$) gene sets and each column represents prioritized ($FDR < 0.05$) genes.

19 Colors on the heatmap represent each gene’s contribution to gene set enrichment (depicted as Z-score,

20 only top 10 highest Z-scores per gene set are visualized). Sidebars represent p-values for GWAS, gene set

21 enrichment (GSE), and gene prioritization (nominal P-value on \log_{10} scale). Top 10 gene sets per

22 annotation category are visualized. GO, Gene Ontology; KE, Kyoto Encyclopedia of Gene and Genomes;

23 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
<i>Loci found in the HapMap GWAS</i>									
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 IKG GWAS

rs75460349	1	27180088	A	0.97	0.086	0.014	4.50×10^{-10}	<i>ZDHHC18</i>
rs1514895	3	170705693	A	0.71	-0.027	0.004	2.70×10^{-09}	<i>EIF5A2</i>
rs112635299	14	94838142	T	0.02	-0.107	0.017	2.10×10^{-10}	<i>SERPINA1/2</i>
rs1189402	15	53728154	A	0.62	0.025	0.004	3.90×10^{-09}	<i>ONECUT1</i>

Additional loci found in the BMI adjusted IKG GWAS

3:47431869	3	47431869	D	0.59	0.024	0.004	1.10×10^{-08}	<i>PTPN23</i>
rs687339	3	135932359	T	0.78	-0.030	0.005	2.80×10^{-10}	<i>MSL2</i>
rs7795281	7	74122854	A	0.76	0.028	0.005	3.10×10^{-08}	<i>GTF2I</i>
rs1736060	8	11664738	T	0.60	0.029	0.004	2.60×10^{-13}	<i>FDFT1</i>
17:58001690	17	58001690	D	0.44	-0.026	0.004	9.50×10^{-10}	<i>RPS6KB1</i>
rs9611441	22	41339367	C	0.49	-0.022	0.004	1.40×10^{-08}	<i>XPNPEP3</i>

25 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
 26 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
 27 are presented, except for the IKG additional findings. For the HapMap loci, the lead variant from the IKG GWAS is presented when the locus was also found in
 28 the IKG GWAS.