

1 **PROTEIN-CODING VARIANTS IMPLICATE NOVEL GENES RELATED TO LIPID HOMEOSTASIS**
2 **CONTRIBUTING TO BODY FAT DISTRIBUTION**

3 Anne E Justice^{¥,1,2}, Tugce Karaderi^{¥,3,4}, Heather M Highland^{¥,1,5}, Kristin L Young^{¥,1}, Mariaelisa Graff^{¥,1},
4 Yingchang Lu^{¥,6,7,8}, Valérie Turcot^{¥,9}, Paul L Auer¹⁰, Rebecca S Fine^{11,12,13}, Xiuqing Guo¹⁴, Claudia
5 Schurmann^{7,8}, Adelheid Lempradl¹⁵, Eirini Marouli¹⁶, Anubha Mahajan³, Thomas W Winkler¹⁷, Adam E
6 Locke^{18,19}, Carolina Medina-Gomez^{20,21}, Tõnu Esko^{11,13,22}, Sailaja Vedantam^{11,12,13}, Ayush Giri²³, Ken Sin
7 Lo^{9,23}, Tamuno Alfred⁷, Poorva Mudgal²⁴, Maggie CY Ng^{24,25}, , Nancy L Heard-Costa^{26,27}, Mary F Feitosa²⁸,
8 Alisa K Manning^{11,29,30}, Sara M Willems³¹, Suthesh Sivapalaratnam^{30,32,33}, , Goncalo Abecasis^{18,34}, Dewan S
9 Alam³⁵, Matthew Allison³⁶, Philippe Amouyel^{37,38,39}, Zorayr Arzumanyan¹⁴, Beverley Balkau⁴⁰, Lisa
10 Bastarache⁴¹, Sven Bergmann^{42,43}, Lawrence F Bielak⁴⁴, Matthias Blüher^{45,46}, Michael Boehnke¹⁸, Heiner
11 Boeing⁴⁷, Eric Boerwinkle^{5,48}, Carsten A Böger⁴⁹, Jette Bork-Jensen⁵⁰, Erwin P Bottinger⁷, Donald W
12 Bowden^{24,25,51}, Ivan Brandslund^{52,53}, Linda Broer²¹, Amber A Burt⁵⁴, Adam S Butterworth^{55,56}, Mark J
13 Caulfield^{16,57}, Giancarlo Cesana⁵⁸, John C Chambers^{59,60,61,62,63}, Daniel I Chasman^{11,64,65,66}, Yii-Der Ida Chen¹⁴,
14 Rajiv Chowdhury⁵⁵, Cramer Christensen⁶⁷, Audrey Y Chu⁶⁵, Francis S Collins⁶⁸, James P Cook⁶⁹, Amanda J
15 Cox^{24,25,70}, David S Crosslin⁷¹, John Danesh^{55,56,72,73}, Paul IW de Bakker^{74,75}, Simon de Denus^{9,76}, Renée de
16 Mutsert⁷⁷, George Dedoussis⁷⁸, Ellen W Demerath⁷⁹, Joe G Dennis⁸⁰, Josh C Denny⁴¹, Emanuele Di
17 Angelantonio^{55,56,73}, Marcus Dörr^{81,82}, Fotios Drenos^{83,84,85}, Marie-Pierre Dubé^{9,86}, Alison M Dunning⁸⁷,
18 Douglas F Easton^{80,87}, Paul Elliott⁸⁸, Evangelos Evangelou^{61,89}, Aliko-Eleni Farmaki⁷⁸, Shuang Feng¹⁸, Ele
19 Ferrannini^{90,91}, Jean Ferrieres⁹², Jose C Florez^{11,29,30}, Myriam Fornage⁹³, Caroline S Fox²⁷, Paul W
20 Franks^{94,95,96}, Nele Friedrich⁹⁷, Wei Gan³, Ilaria Gandin⁹⁸, Paolo Gasparini^{99,100}, Vilmantas Giedraitis¹⁰¹,
21 Giorgia Grotto^{99,100}, Mathias Gorski^{17,49}, Harald Grallert^{102,103,104}, Niels Grarup⁵⁰, Megan L Grove⁵, Stefan
22 Gustafsson¹⁰⁵, Jeff Haessler¹⁰⁶, Torben Hansen⁵⁰, Andrew T Hattersley¹⁰⁷, Caroline Hayward¹⁰⁸, Iris M
23 Heid^{17,109}, Oddgeir L Holmen¹¹⁰, G Kees Hovingh¹¹¹, Joanna MM Howson⁵⁵, Yao Hu¹¹², Yi-Jen Hung^{113,114},

24 Kristian Hveem^{110,115}, M Arfan Ikram^{20,116,117}, Erik Ingelsson^{105,118}, Anne U Jackson¹⁸, Gail P Jarvik^{54,119},
25 Yucheng Jia¹⁴, Torben Jørgensen^{120,121,122}, Pekka Jousilahti¹²³, Johanne M Justesen⁵⁰, Bratati
26 Kahali^{124,125,126,127}, Maria Karaleftheri¹²⁸, Sharon LR Kardia⁴⁴, Fredrik Karpe^{129,130}, Frank Kee¹³¹, Hidetoshi
27 Kitajima³, Pirjo Komulainen¹³², Jaspal S Kooner^{60,62,63,133}, Peter Kovacs⁴⁵, Bernhard K Krämer¹³⁴, Kari
28 Kuulasmaa¹²³, Johanna Kuusisto¹³⁵, Markku Laakso¹³⁵, Timo A Lakka^{132,136,137}, David Lamparter^{42,43,138}, Leslie
29 A Lange¹³⁹, Claudia Langenberg³¹, Eric B Larson^{54,140,141}, Nanette R Lee^{142,143}, Wen-Jane Lee^{144,145}, Terho
30 Lehtimäki^{146,147}, Cora E Lewis¹⁴⁸, Huaixing Li¹¹², Jin Li¹⁴⁹, Ruifang Li-Gao⁷⁷, Li-An Lin⁹³, Xu Lin¹¹², Lars Lind¹⁵⁰,
31 Jaana Lindström¹²³, Allan Linneberg^{122,151,152}, Ching-Ti Liu¹⁵³, Dajiang J Liu¹⁵⁴, Jian'an Luan³¹, Leo-Pekka
32 Lyytikäinen^{146,147}, Stuart MacGregor¹⁵⁵, Reedik Mägi²², Satu Männistö¹²³, Gaëlle Marenne⁷², Jonathan
33 Marten¹⁰⁸, Nicholas GD Masca^{156,157}, Mark I McCarthy^{3,129,130}, Karina Meidtner^{102,158}, Evelin Mihailov²²,
34 Leena Moilanen¹⁵⁹, Marie Moitry^{160,161}, Dennis O Mook-Kanamori^{77,162}, Anna Morgan⁹⁹, Andrew P
35 Morris^{3,69}, Martina Müller-Nurasyid^{109,163,164}, Patricia B Munroe^{16,57}, Narisu Narisu⁶⁸, Christopher P
36 Nelson^{156,157}, Matt Neville^{129,130}, Ioanna Ntalla¹⁶, Jeffrey R O'Connell¹⁶⁵, Katharine R Owen^{129,130}, Oluf
37 Pedersen⁵⁰, Gina M Peloso¹⁵³, Craig E Pennell^{166,167}, Markus Perola^{123,168}, James A Perry¹⁶⁵, John RB Perry³¹,
38 Tune H Pers^{50,169}, Ailith Ewing⁸⁰, Ozren Polasek^{170,171}, Olli T Raitakari^{172,173}, Asif Rasheed¹⁷⁴, Chelsea K
39 Raulerson¹⁷⁵, Rainer Rauramaa^{132,136}, Dermot F Reilly¹⁷⁶, Alex P Reiner^{106,177}, Paul M Ridker^{65,66,178}, Manuel
40 A Rivas¹⁷⁹, Neil R Robertson^{3,129}, Antonietta Robino¹⁸⁰, Igor Rudan¹⁷¹, Katherine S Ruth¹⁸¹, Danish
41 Saleheen^{174,182}, Veikko Salomaa¹²³, Nilesh J Samani^{156,157}, Pamela J Schreiner¹⁸³, Matthias B Schulze^{102,158},
42 Robert A Scott³¹, Marcelo Segura-Lepe⁶¹, Xueling Sim^{18,184}, Andrew J Slater^{185,186}, Kerrin S Small¹⁸⁷, Blair H
43 Smith^{188,189}, Jennifer A Smith⁴⁴, Lorraine Southam^{3,72}, Timothy D Spector¹⁸⁷, Elizabeth K Speliotes^{124,125,126},
44 Kari Stefansson^{190,191}, Valgerdur Steinthorsdottir¹⁹⁰, Kathleen E Stirrups^{16,33}, Konstantin Strauch^{109,192},
45 Heather M Stringham¹⁸, Michael Stumvoll^{45,46}, Liang Sun¹¹², Praveen Surendran⁵⁵, Karin MA Swart¹⁹³, Jean-
46 Claude Tardif^{9,86}, Kent D Taylor¹⁴, Alexander Teumer¹⁹⁴, Deborah J Thompson⁸⁰, Gudmar Thorleifsson¹⁹⁰,
47 Unnur Thorsteinsdottir^{190,191}, Betina H Thuesen¹²², Anke Tönjes¹⁹⁵, Mina Torres¹⁹⁶, Emmanouil

48 Tsafantakis¹⁹⁷, Jaakko Tuomilehto^{123,198,199,200}, André G Uitterlinden^{20,21}, Matti Uusitupa²⁰¹, Cornelia M van
49 Duijn²⁰, Mauno Vanhala^{202,203}, Rohit Varma¹⁹⁶, Sita H Vermeulen²⁰⁴, Henrik Vestergaard^{50,205}, Veronique
50 Vitart¹⁰⁸, Thomas F Vogt²⁰⁶, Dragana Vuckovic^{99,100}, Lynne E Wagenknecht²⁰⁷, Mark Walker²⁰⁸, Lars
51 Wallentin²⁰⁹, Feijie Wang¹¹², Carol A Wang^{166,167}, Shuai Wang¹⁵³, Nicholas J Wareham³¹, Helen R
52 Warren^{16,57}, Dawn M Waterworth²¹⁰, Jennifer Wessel²¹¹, Harvey D White²¹², Cristen J Willer^{124,125,213}, James
53 G Wilson²¹⁴, Andrew R Wood¹⁸¹, Ying Wu¹⁷⁵, Hanieh Yaghoobkar¹⁸¹, Jie Yao¹⁴, Laura M Yerges-
54 Armstrong^{165,215}, Robin Young^{55,216}, Eleftheria Zeggini⁷², Xiaowei Zhan²¹⁷, Weihua Zhang^{60,61}, Jing Hua
55 Zhao³¹, Wei Zhao¹⁸², He Zheng¹¹², Wei Zhou^{124,125}, M Carola Zillikens^{20,21}, CHD Exome+ Consortium, Cohorts
56 for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, EPIC-CVD Consortium,
57 ExomeBP Consortium, Global Lipids Genetic Consortium, GoT2D Genes Consortium, InterAct, ReproGen
58 Consortium, T2D-Genes Consortium, The MAGIC Investigators, Fernando Rivadeneira^{20,21}, Ingrid B
59 Borecki²⁸, J. Andrew Pospisilik¹⁵, Panos Deloukas^{16,218}, Timothy M Frayling¹⁸¹, Guillaume Lettre^{9,86}, Karen L
60 Mohlke¹⁷⁵, Jerome I Rotter¹⁴, Zoltán Kutalik^{43,219}, Joel N Hirschhorn^{11,13,220}, L Adrienne Cupples^{1,27,153}, Ruth
61 JF Loos^{1,7,8,221}, Kari E North^{1,222}, Cecilia M Lindgren^{1,*,3,223}

62

63 ¥ These authors contributed equally to this work.

64 £ These authors jointly supervised this work.

65 *CORRESPONDING AUTHORS

66 Prof. Kari North

67 Department of Epidemiology

68 University of North Carolina at Chapel Hill

69 137 East Franklin Street

70 Suite 306

71 Chapel Hill, NC 27514
72
73 Prof. Cecilia M Lindgren
74 The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery
75 University of Oxford
76 Roosevelt Drive
77 Oxford
78 OX3 7BN
79 United Kingdom
80 celi@well.ox.ac.uk

81 **AFFILIATIONS**

- 82 1. Department of Epidemiology, University of North Carolina, Chapel Hill, NC, 27514, USA
- 83 2. Weis Center for Research, Geisinger Health System, Danville, PA 17822
- 84 3. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
- 85 4. Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean
86 University, Famagusta, Cyprus
- 87 5. Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental
88 Sciences, School of Public Health, The University of Texas Health Science Center at Houston,
89 Houston, TX, 77030, USA
- 90 6. Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt
91 Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, 37203, USA
- 92 7. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount
93 Sinai, New York, NY, 10029, USA

- 94 8. The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount
95 Sinai, New York, NY, 10069, USA
- 96 9. Montreal Heart Institute, Universite de Montreal, Montreal, Quebec, H1T 1C8, Canada
- 97 10. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, 53201, USA
- 98 11. Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
- 99 12. Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA
- 100 13. Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston
101 Children's Hospital, Boston, MA, 02115, USA
- 102 14. Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical
103 Center, Torrance, CA, 90502, USA
- 104 15. Max Planck Institute of Immunobiology and Epigenetics, Freiburg, 79108, Germany
- 105 16. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen
106 Mary University of London, London, EC1M 6BQ, UK
- 107 17. Department of Genetic Epidemiology, University of Regensburg, Regensburg, D-93051, Germany
- 108 18. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor,
109 MI, 48109, USA
- 110 19. McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, 63108,
111 USA
- 112 20. Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- 113 21. Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- 114 22. Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia
- 115 23. Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt
116 Genetics Institute, Vanderbilt University, Nashville, TN, 37203, USA
- 117 24. Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA

- 118 25. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine,
119 Winston-Salem, NC, 27157, USA
- 120 26. Department of Neurology, Boston University School of Medicine, Boston, MA, 02118, USA
- 121 27. NHLBI Framingham Heart Study, Framingham, MA, 01702, USA
- 122 28. Division of Statistical Genomics, Department of Genetics, Washington University School of
123 Medicine, St. Louis, MO, 63108, USA
- 124 29. Department of Medicine, Harvard University Medical School, Boston, MA, 02115, USA
- 125 30. Massachusetts General Hospital, Boston, MA, 02114, USA
- 126 31. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of
127 Metabolic Science, Cambridge, CB2 0QQ, UK
- 128 32. Department of Vascular Medicine, AMC, Amsterdam, 1105 AZ, The Netherlands
- 129 33. Department of Haematology, University of Cambridge, Cambridge, CB2 0PT, UK
- 130 34. School of Public Health, University of Michigan, Ann Arbor, MI, 48109, USA
- 131 35. School of Kinesiology and Health Science, Faculty of Health, York University, Toronto
- 132 36. Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA,
133 92093, USA
- 134 37. INSERM U1167, Lille, F-59019, France
- 135 38. Institut Pasteur de Lille, U1167, Lille, F-59019, France
- 136 39. Universite de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related
137 diseases, Lille, F-59019, France
- 138 40. INSERM U1018, Centre de recherche en Épidémiologie et Sante des Populations (CESP), Villejuif,
139 France
- 140 41. Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, 37203, USA
- 141 42. Department of Computational Biology, University of Lausanne, Lausanne, 1011, Switzerland

- 142 43. Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
- 143 44. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI,
144 48109, USA
- 145 45. IFB Adiposity Diseases, University of Leipzig, Leipzig, 04103, Germany
- 146 46. University of Leipzig, Department of Medicine, Leipzig, 04103, Germany
- 147 47. Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE),
148 Nuthetal, 14558, Germany
- 149 48. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030 USA
- 150 49. Department of Nephrology, University Hospital Regensburg, Regensburg, 93042, Germany
- 151 50. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical
152 Sciences, University of Copenhagen, Copenhagen, 2100, Denmark
- 153 51. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
- 154 52. Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, 7100, Denmark
- 155 53. Institute of Regional Health Research, University of Southern Denmark, Odense, 5000, Denmark
- 156 54. Department of Medicine (Medical Genetics), University of Washington, Seattle, WA, 98195, USA
- 157 55. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care,
158 University of Cambridge, Cambridge, CB1 8RN, UK
- 159 56. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public
160 Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
- 161 57. NIHR Barts Cardiovascular Research Centre, Barts and The London School of Medicine & Dentistry,
162 162 Queen Mary University of London, London, EC1M 6BQ, UK
- 163 58. Research Centre on Public Health, University of Milano-Bicocca, Monza, 20900, Italy
- 164 59. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232,
165 Singapore

- 166 60. Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, Middlesex,
167 UB1 3HW, UK
- 168 61. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,
169 London, W2 1PG, UK
- 170 62. Imperial College Healthcare NHS Trust, London, W12 0HS, UK
- 171 63. MRC-PHE Centre for Environment and Health, Imperial College London, London, W2 1PG, UK
- 172 64. Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA,
173 02115, USA
- 174 65. Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston, MA,
175 02215, USA
- 176 66. Harvard Medical School, Boston, MA, 02115, USA
- 177 67. Medical department, Lillebaelt Hospital, Vejle, 7100, Denmark
- 178 68. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute,
179 National Institutes of Health, Bethesda, MD, 20892, USA
- 180 69. Department of Biostatistics, University of Liverpool, Liverpool, L69 3GL, UK
- 181 70. Menzies Health Institute Queensland, Griffith University, Southport, QLD, Australia
- 182 71. Department of Biomedical Informatics and Medical Education, University of Washington, Seattle,
183 WA, 98195, USA
- 184 72. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK
- 185 73. British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of
186 Cambridge, Cambridge, CB2 0QQ, UK
- 187 74. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht,
188 The Netherlands

- 189 75. Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht,
190 Utrecht, 3584 CX, The Netherlands
- 191 76. Faculty of Pharmacy, Universite de Montreal, Montreal, Quebec, H3T 1J4, Canada
- 192 77. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2300RC, The
193 Netherlands
- 194 78. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio
195 University, Athens, 17671, Greece
- 196 79. Division of Epidemiology & Community Health, School of Public Health, University of Minnesota,
197 Minneapolis, MN, 55454, USA
- 198 80. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care,
199 University of Cambridge, Cambridge, CB1 8RN, UK
- 200 81. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, 17475, Germany
- 201 82. DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, 17475,
202 Germany
- 203 83. Institute of Cardiovascular Science, University College London, London, WC1E 6JF, UK
- 204 84. MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of
205 Bristol, Bristol, BS8 2BN, UK
- 206 85. Department of Life Sciences, Brunel University London, Uxbridge, UB8 3PH, UK
- 207 86. Department of Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, H3T
208 1J4, Canada
- 209 87. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge,
210 Cambridge, CB1 8RN, UK
- 211 88. Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health,
212 School of Public Health, Imperial College London, London, W2 1PG, UK

- 213 89. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina,
214 45110, Greece
- 215 90. CNR Institute of Clinical Physiology, Pisa, Italy
- 216 91. Department of Clinical & Experimental Medicine, University of Pisa, Italy
- 217 92. Toulouse University School of Medicine, Toulouse, TSA 50032 31059, France
- 218 93. Institute of Molecular Medicine, The University of Texas Health Science Center at Houston,
219 Houston, TX, 77030, USA
- 220 94. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University,
221 Malmo, SE-20502, Sweden
- 222 95. Department of Nutrition, Harvard School of Public Health, Boston, MA, 02115, USA
- 223 96. Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, 901
224 87, Sweden
- 225 97. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald,
226 Greifswald, 17475, Germany
- 227 98. Ilaria Gandin, Research Unit, AREA Science Park, Trieste, 34149, Italy
- 228 99. Department of Medical Sciences, University of Trieste, Trieste, 34137, Italy
- 229 100. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy
- 230 101. Geriatrics, Department of Public Health, Uppsala University, Uppsala, 751 85, Sweden
- 231 102. German Center for Diabetes Research, München-Neuherberg, 85764, Germany
232
- 233 103. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for
234 Environmental Health, Neuherberg, 85764, Germany
- 235 104. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research
236 Center for Environmental Health, Neuherberg, 85764, Germany

- 237 105. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory,
238 Uppsala University, Uppsala, 751 41, Sweden
- 239 106. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, 98109,
240 USA
- 241 107. University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
- 242 108. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh,
243 Edinburgh, EH4 2XU, UK
- 244 109. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for
245 Environmental Health, Neuherberg, 85764, Germany
- 246 110. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian
247 University of Science and Technology, Trondheim, 7600, Norway
- 248 111. AMC, Department of Vascular Medicine, Amsterdam, 1105 AZ, The Netherlands
- 249 112. CAS Key Laboratory of Nutrition, Metabolism and Food safety, Shanghai Institute of Nutrition and
250 Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences,
251 Chinese Academy of Sciences, Shanghai 200031, China
- 252 113. Division of Endocrinology and Metabolism, Department of Internal Medicine, Tri-Service General
253 Hospital Songshan Branch, Taipei, Taiwan 11
- 254 114. School of Medicine, National Defense Medical Center, Taipei, Taiwan 114, Taiwan
- 255 115. HUNT Research center, Department of Public Health, Norwegian University of Science and
256 Technology, Levanger, 7600, Norway
- 257 116. Department of Neurology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- 258 117. Department of Radiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- 259 118. Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305, USA
- 260 119. Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA

261 120. Faculty of medicine, Aalborg University, Aalborg, DK-9000, Denmark

262 121. Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen,
263 Copenhagen, 2200, Denmark

264 122. Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, DK-2600,
265 Denmark

266 123. National Institute for Health and Welfare, Helsinki, FI-00271, Finland

267 124. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor,
268 MI, 48109, USA

269 125. Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109, USA

270 126. Division of Gastroenterology, University of Michigan, Ann Arbor, MI, 48109, USA

271 127. Centre for Brain Research, Indian Institute of Science, Bangalore 560012, India

272 128. Echinops Medical Centre, Echinops, Greece

273 129. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine,
274 University of Oxford, Oxford, OX3 7LE, UK

275 130. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK

276 131. UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast, UK,
277 BT12 6BJ, UK

278 132. Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise
279 Medicine, Kuopio, 70100, Finland

280 133. National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus,
281 London, W12 0NN, UK

282 134. University Medical Centre Mannheim, 5th Medical Department, University of Heidelberg,
283 Mannheim, 68167, Germany

- 284 135. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio
285 University Hospital, Kuopio, 70210, Finland
- 286 136. Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus,
287 70210, Finland
- 288 137. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio,
289 Finland
- 290 138. Verge Genomics, San Fransico, CA, USA
- 291 139. Division of Biomedical and Personalized Medicine, Department of Medicine, University of
292 Colorado-Denver, Aurora, CO, 80045, USA
- 293 140. Kaiser Permanente Washington Health Research Institute Seattle WA 98101
- 294 141. Department of Health Services, University of Washington, Seattle WA 98101
- 295 142. Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, 6000,
296 Philippines
- 297 143. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, 6000,
298 Philippines
- 299 144. Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan 407,
300 Taiwan
- 301 145. Department of Social Work, Tunghai University, Taichung, Taiwan
- 302 146. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, 33521, Finland
- 303 147. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of
304 Medicine and Life Sciences, University of Tampere, Tampere 33014, Finland
- 305 148. Division of Preventive Medicine University of Alabama at Birmingham, Birmingham, AL 35205,
306 USA

- 307 149. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of
308 Medicine, Palo Alto, CA, 94304, USA
- 309 150. Uppsala University, Uppsala, 75185, Sweden
- 310 151. Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, DK-2000,
311 Frederiksberg, Denmark
- 312 152. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of
313 Copenhagen, Copenhagen, 2200, Denmark
- 314 153. Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA
- 315 154. Department of Public Health Sciences, Institute for Personalized Medicine, the Pennsylvania State
316 University College of Medicine, Hershey, PA, 17033, USA
- 317 155. QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
- 318 156. Department of Cardiovascular Sciences, Univeristy of Leicester, Glenfield Hospital, Leicester, LE3
319 9QP, UK
- 320 157. NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP,
321 UK
- 322 158. Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-
323 Rehbruecke (DIfE), Nuthetal, 14558, Germany
- 324 159. Department of Medicine, Kuopio University Hospital, Kuopio, 70210, Finland
- 325 160. Department of Epidemiology and Public Health, University of Strasbourg, Strasbourg, F-67085,
326 France
- 327 161. Department of Public Health, University Hospital of Strasbourg, Strasbourg, F-67091, France
- 328 162. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, 2300RC,
329 The Netherlands

- 330 163. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universitat,
331 Munich, 81377, Germany
- 332 164. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich,
333 80802, Germany
- 334 165. Program for Personalized and Genomic Medicine, Department of Medicine, University of
335 Maryland School of Medicine, Baltimore, MD, 21201, US
- 336 166. Division of Obstetric and Gynaecology, School of Medicine, The University of Western Australia,
337 Perth, Western Australia, 6009, Australia
- 338 167. School of Medicine and Public Health, Faculty of Medicine and Health, The University of
339 Newcastle, Newcastle, New South Wales, 2308, Australia
- 340 168. University of Helsinki, Institute for Molecular Medicine (FIMM) and Diabetes and Obesity
341 Research Program, Helsinki, FI00014, Finland
- 342 169. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, 2200, Denmark
- 343 170. School of Medicine, University of Split, Split, 21000, Croatia
- 344 171. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics,
345 University of Edinburgh, Edinburgh, EH8 9AG, UK
- 346 172. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, 20521,
347 Finland
- 348 173. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku,
349 20520, Finland
- 350 174. Centre for Non-Communicable Diseases, Karachi, Pakistan
- 351 175. Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA
- 352 176. Merck & Co., Inc., Genetics and Pharmacogenomics, Boston, MA, 02115, USA
- 353 177. Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA

- 354 178. Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School,
355 Boston, MA, 02115, USA
- 356 179. Department of Biomedical Data Science, Stanford University, Stanford, California 94305
- 357 180. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, 34137, Italy
- 358 181. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2
359 5DW, UK
- 360 182. Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of
361 Pennsylvania, Philadelphia, PA, 19104, USA
- 362 183. Division of Epidemiology & Community Health University of Minnesota, Minneapolis, MN, 55454,
363 USA
- 364 184. Saw Swee Hock School of Public Health, National University Health System, National University of
365 Singapore, Singapore 117549, Singapore
- 366 185. Genetics, Target Sciences, GlaxoSmithKline, Research Triangle Park, NC, 27709, US
- 367 186. OmicSoft a QIAGEN Company, Cary, NC, 27513, US
- 368 187. Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH,
369 UK
- 370 188. Division of Population Health Sciences, Ninewells Hospital and Medical School, University of
371 Dundee, Dundee, UK
- 372 189. Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh,
373 Edinburgh, EH4 2XU, UK
- 374 190. deCODE Genetics/Amgen inc., Reykjavik, 101, Iceland
- 375 191. Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland
- 376 192. Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, 81377, Germany

377 193. VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam, 1007
378 MB, The Netherlands

379 194. Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany

380 195. Center for Pediatric Research, Department for Women's and Child Health, University of Leipzig,
381 Leipzig, 04103, Germany

382 196. USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University
383 of Southern California, Los Angeles, CA, 90033, USA

384 197. Anogia Medical Centre, Anogia, Greece

385 198. Centre for Vascular Prevention, Danube-University Krems, Krems, 3500, Austria

386 199. Dasman Diabetes Institute, Dasman, 15462, Kuwait

387 200. Diabetes Research Group, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

388 201. Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, 70210,
389 Finland

390 202. Central Finland Central Hospital, Jyvaskyla, 40620, Finland

391 203. University of Eastern Finland, Kuopio, 70210, Finland

392 204. Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, 6500 HB,
393 The Netherlands

394 205. Steno Diabetes Center Copenhagen, Gentofte, 2800, Denmark

395 206. Merck & Co., Inc., Cardiometabolic Disease, Kenilworth, NJ, 07033, USA

396 207. Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, 27157,
397 USA

398 208. Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, NE2 4HH, UK

399 209. Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala
400 University, Uppsala, 752 37, Sweden

401 210. Genetics, Target Sciences, GlaxoSmithKline, Collegeville, PA

402 211. Departments of Epidemiology & Medicine, Diabetes Translational Research Center, Fairbanks
403 School of Public Health & School of Medicine, Indiana University, Indiana, IN, 46202, USA

404 212. Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland,
405 New Zealand

406 213. Department of Human Genetics, University of Michigan, Ann Arbor, MI, 48109, USA

407 214. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS,
408 39216, USA

409 215. GlaxoSmithKline, King of Prussia, PA, 19406, USA

410 216. University of Glasgow, Glasgow, G12 8QQ, UK

411 217. Department of Clinical Sciences, Quantitative Biomedical Research Center, Center for the
412 Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, 75390,
413 USA

414 218. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-
415 HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia

416 219. Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, 1010,
417 Switzerland

418 220. Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, 02115, USA

419 221. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai,
420 New York, NY, 10069, USA

421 222. Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, NC, 27514,
422 USA

423 223. Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of
424 Oxford, Oxford, OX3 7BN, UK

425

426

427 **ABSTRACT**

428 Body fat distribution is a heritable risk factor for a range of adverse health consequences,
429 including hyperlipidemia and type 2 diabetes. To identify protein-coding variants associated with body fat
430 distribution, assessed by waist-to-hip ratio adjusted for body mass index, we analyzed 228,985 predicted
431 coding and splice site variants available on exome arrays in up to 344,369 individuals from five major
432 ancestries for discovery and 132,177 independent European-ancestry individuals for validation. We
433 identified 15 common (minor allele frequency, MAF \geq 5%) and 9 low frequency or rare (MAF < 5%) coding
434 variants that have not been reported previously. Pathway/gene set enrichment analyses of all associated
435 variants highlight lipid particle, adiponectin level, abnormal white adipose tissue physiology, and bone
436 development and morphology as processes affecting fat distribution and body shape. Furthermore, the
437 cross-trait associations and the analyses of variant and gene function highlight a strong connection to
438 lipids, cardiovascular traits, and type 2 diabetes. In functional follow-up analyses, specifically in *Drosophila*
439 RNAi-knockdown crosses, we observed a significant increase in the total body triglyceride levels for two
440 genes (*DNAH10* and *PLXND1*). By examining variants often poorly tagged or entirely missed by genome-
441 wide association studies, we implicate novel genes in fat distribution, stressing the importance of
442 interrogating low-frequency and protein-coding variants.

443

444

445

446

447

448

449 Body fat distribution, as assessed by waist-to-hip ratio (WHR), is a heritable trait and a well-
450 established risk factor for adverse metabolic outcomes¹⁻⁶. A high WHR often indicates a large presence
451 of intra-abdominal fat whereas a low WHR is correlated with a greater accumulation of gluteofemoral
452 fat. Lower values of WHR have been consistently associated with lower risk of cardiometabolic diseases
453 like type 2 diabetes (T2D)^{7,8}, or differences in bone structure and gluteal muscle mass⁹. These
454 epidemiological associations are consistent with the results of our previously reported genome-wide
455 association study (GWAS) of 49 loci associated with WHR (after adjusting for body mass index,
456 WHRadjBMI)¹⁰. Notably, a genetic predisposition to higher WHRadjBMI is associated with increased risk
457 of T2D and coronary heart disease (CHD), and this association appears to be causal⁹.

458 More recently, large-scale genetic studies have identified ~125 common loci for central obesity,
459 primarily non-coding variants of relatively modest effect, for different measures of body fat distribution¹⁰⁻
460 ¹⁶. Large scale interrogation of both common (minor allele frequency [MAF]≥5%) and low frequency or
461 rare (MAF<5%) coding and splice site variation may lead to additional insights into the genetic and
462 biological etiology of central obesity by narrowing in on causal genes contributing to trait variance. Thus,
463 we set out to identify protein-coding and splice site variants associated with WHRadjBMI using exome
464 array data and to explore their contribution to variation in WHRadjBMI through multiple follow-up
465 analyses.

466 **RESULTS**

467 **Protein-coding and splice site variation associated with body fat distribution**

468 We conducted a 2-stage fixed-effects meta-analysis testing both additive and recessive models in
469 order to detect protein-coding genetic variants that influence WHRadjBMI (**Online Methods, Figure 1**).
470 Our stage 1 meta-analysis included up to 228,985 variants (218,195 with MAF<5%) in up to 344,369
471 individuals from 74 studies of European (N=288,492), South Asian (N=29,315), African (N=15,687), East

472 Asian (N=6,800) and Hispanic/Latino (N=4,075) descent, genotyped with an ExomeChip array
473 **(Supplementary Tables 1-3)**. For stage 2, we assessed 70 suggestively significant ($P < 2 \times 10^{-6}$) variants from
474 stage 1 in two independent cohorts from the United Kingdom [UK Biobank (UKBB), N=119,572] and
475 Iceland (deCODE, N=12,605) **(Online Methods, Supplementary Data 1-3)** for a total stage 1+2 sample size
476 of 476,546 (88% European). Variants were considered statistically significant in the total meta-analyzed
477 sample (stage 1+2) when they achieved a significance threshold of $P < 2 \times 10^{-7}$ after Bonferroni correction
478 for multiple testing (0.05/246,328 variants tested). Of the 70 variants brought forward, two common and
479 five rare variants were not available in either Stage 2 study **(Tables 1-2, Supplementary Data 1-3)**. Thus,
480 we require $P < 2 \times 10^{-7}$ in Stage 1 for significance. Variants are considered novel if they were greater than
481 one megabase (Mb) from a previously-identified WHRadjBMI lead SNP¹⁰⁻¹⁶.

482 In stages 1 and 2 combined all ancestry meta-analyses, we identified 48 coding variants (16 novel)
483 across 43 genes, 47 identified assuming an additive model, and one more variant under a recessive model
484 **(Table 1, Supplementary Figures 1-4)**. Due to the possible heterogeneity introduced by combining
485 multiple ancestries¹⁷, we also performed a European-only meta-analysis. Here, four additional coding
486 variants were significant (three novel) assuming an additive model **(Table 1, Supplementary Figures 5-8)**.
487 Of these 52 significant variants (48 from the all ancestry and 4 from the European-only analyses), eleven
488 were of low frequency, including seven novel variants in *RAPGEF3*, *FGFR2*, *R3HDML*, *HIST1H1T*, *PCNXL3*,
489 *ACVR1C*, and *DARS2*. These low frequency variants tended to display larger effect estimates than any of
490 the previously reported common variants **(Figure 2)**¹⁰. In general, variants with MAF < 1% had effect sizes
491 approximately three times greater than those of common variants (MAF > 5%). Although, we cannot rule
492 out the possibility that additional rare variants with smaller effects sizes exist that, despite our ample
493 sample size, we are still underpowered to detect (See estimated 80% power in **Figure 2**). However, in the
494 absence of common variants with similarly large effects, our results point to the importance of
495 investigating rare and low frequency variants to identify variants with large effects **(Figure 2)**.

496 Given the established differences in the genetic underpinnings between sexes for
497 WHRadjBMI^{10,11}, we also performed sex-stratified analyses and report variants that were array-wide
498 significant ($P < 2 \times 10^{-7}$) in at least one sex stratum and exhibit significant sex-specific effects ($P_{\text{sexhet}} < 7.14 \times 10^{-4}$, see **Online Methods**). We found four additional novel variants that were not identified in the sex-
500 combined meta-analyses (in *UGGT2* and *MMP14* for men only; and *DSTYK* and *ANGPTL4* for women only)
501 (**Table 2, Supplementary Figures 9-15**). Variants in *UGGT2* and *ANGPTL4* were of low frequency
502 ($\text{MAF}_{\text{men}} = 0.6\%$ and $\text{MAF}_{\text{women}} = 1.9\%$, respectively). Additionally, 14 variants from the sex-combined meta-
503 analyses displayed stronger effects in women, including the novel, low frequency variant in *ACVR1C*
504 (rs55920843, $\text{MAF} = 1.1\%$, **Supplementary Figure 4**). Overall, 19 of the 56 variants (32%) identified across
505 all meta-analyses (48 from all ancestry, 4 from European-only and 4 from sex-stratified analyses) showed
506 significant sex-specific effects on WHRadjBMI (**Figure 1**): 16 variants with significantly stronger effects in
507 women, and three in men (**Figure 1**).

508 In summary, we identified 56 array-wide significant coding variants ($P < 2.0 \times 10^{-7}$); 43 common (14
509 novel) and 13 low frequency or rare variants (9 novel). For all 55 significant variants from the additive
510 model (47 from all ancestry, 4 from European-only, and 4 from sex-specific analyses), we examined
511 potential collider bias^{18,19}, i.e. potential bias in effect estimates caused by adjusting for a correlated and
512 heritable covariate like BMI, for the relevant sex stratum and ancestry. We corrected each of the variant
513 - WHRadjBMI associations for the correlation between WHR and BMI and the correlation between the
514 variant and BMI (**Online Methods, Supplementary Table 7, Supplementary Note 1**). Overall, 51 of the 55
515 additive model variants were robust against collider bias^{18,19} across all primary and secondary meta-
516 analyses. Of the 55, 25 of the WHRadjBMI variants from the additive model were nominally associated
517 with BMI ($P_{\text{BMI}} < 0.05$), yet effect sizes changed little after correction for potential biases (15% change in
518 effect estimate on average). For 4 of the 55 SNPs (rs141845046, rs1034405, rs3617, rs9469913, **Table 1**),
519 the association with WHRadjBMI appears to be attenuated following correction ($P_{\text{corrected}} > 9 \times 10^{-4}$,

520 0.05/55), including one novel variant, rs1034405 in *C3orf18*. Thus, these 4 variants warrant further
521 functional investigations to quantify their impact on WHR, as a true association may still exist, although
522 the effect may be slightly overestimated in the current analysis.

523 Using stage 1 meta-analysis results, we then aggregated low frequency variants across genes and
524 tested their joint effect with both SKAT and burden tests²⁰ (**Supplementary Table 8, Online Methods**). We
525 identified five genes that reached array-wide significance ($P < 2.5 \times 10^{-6}$, 0.05/16,222 genes tested),
526 *RAPGEF3*, *ACVR1C*, *ANGPTL4*, *DNAI1*, and *NOP2*. However, while all genes analyzed included more than
527 one variant, none remained significant after conditioning on the single variant with the most significant
528 p-value. We identified variants within *RAPGEF3*, *ACVR1C*, *ANGPTL4* that reached suggestive significance
529 in Stage 1 and chip-wide significance in stage 1+2 for one or more meta-analyses (**Tables 1 and 2**);
530 however, we did not identify any significant variants for *DNAI1* and *NOP2*. While neither of these genes
531 had a single variant that reached chip-wide significance, they each had variants with nearly significant
532 results (*NOP2*: $P = 3.69 \times 10^{-5}$, *DNAI1*: 4.64×10^{-5}). Combined effects with these single variants and others in
533 LD within the gene likely drove the association in our aggregate gene-based tests, but resulted in non-
534 significance following conditioning on the top variant. While our results suggest these associations are
535 driven by a single variant, each gene may warrant consideration in future investigations.

536

537 **Conditional analyses**

538 We next implemented conditional analyses to determine (1) the number of independent
539 association signals the 56 array-wide significant coding variants represent, and (2) whether the 33 variants
540 near known GWAS association signals (± 1 Mb) represent independent novel association signals. To
541 determine if these variants were independent association signals, we used approximate joint conditional
542 analyses to test for independence in stage 1 (**Online Methods; Supplementary Table 4**)²⁰. Only the *RSPO3*-
543 *KIAA0408* locus contains two independent variants 291 Kb apart, rs1892172 in *RSPO3* (MAF=46.1%,

544 $P_{\text{conditional}}=4.37 \times 10^{-23}$ in the combined sexes, and $P_{\text{conditional}}=2.4 \times 10^{-20}$ in women) and rs139745911 in
545 *KIAA0408* (MAF=0.9%, $P_{\text{conditional}}=3.68 \times 10^{-11}$ in the combined sexes, and $P_{\text{conditional}}=1.46 \times 10^{-11}$ in women;
546 **Figure 3A**).

547 Further, 33 of our significant variants are within one Mb of previously identified GWAS tag SNPs
548 for WHRadjBMI. We again used approximate joint conditional analysis to test for independence in the
549 stage 1 meta-analysis dataset and obtained further complementary evidence from the UKBB dataset
550 where necessary (**Online Methods**). We identified one coding variant representing a novel independent
551 signal in a known locus [*RREB1*; stage1 meta-analysis, rs1334576, EAF = 0.44, $P_{\text{conditional}}= 3.06 \times 10^{-7}$,
552 (**Supplementary Table 5, Figure 3 [B]**); UKBB analysis, rs1334576, *RREB1*, $P_{\text{conditional}}= 1.24 \times 10^{-8}$,
553 (**Supplementary Table 6**) in the sex-combined analysis.

554 In summary, we identified a total of 56 WHRadjBMI-associated coding variants in 41 independent
555 association signals. Of these 41 independent association signals, 24 are new or independent of known
556 GWAS-identified tag SNPs (either >1MB +/- or array-wide significant following conditional analyses)
557 (**Figure 1**). Thus, bringing our total to 15 common and 9 low-frequency or rare novel variants following
558 conditional analyses. The remaining non-GWAS-independent variants may assist in narrowing in on the
559 causal variant or gene underlying these established association signals.

560 **Gene set and pathway enrichment analysis**

561 To determine if the significant coding variants highlight novel biological pathways and/or provide
562 additional support for previously identified biological pathways, we applied two complementary pathway
563 analysis methods using the EC-DEPICT (ExomeChip Data-driven Expression Prioritized Integration for
564 Complex Traits) pathway analysis tool,^{21,22} and PASCAL²³ (**Online Methods**). While for PASCAL all variants
565 were used, in the case of EC-DEPICT, we examined 361 variants with suggestive significance ($P < 5 \times 10^{-4}$)^{10,17}
566 from the combined ancestries and combined sexes analysis (which after clumping and filtering became

567 101 lead variants in 101 genes). We separately analyzed variants that exhibited significant sex-specific
568 effects ($P_{\text{sexhet}} < 5 \times 10^{-4}$).

569 The sex-combined analyses identified 49 significantly enriched gene sets ($\text{FDR} < 0.05$) that grouped
570 into 25 meta-gene sets (**Supplementary Note 2, Supplementary Data 4-5**). We noted a cluster of meta-
571 gene sets with direct relevance to metabolic aspects of obesity (“enhanced lipolysis,” “abnormal glucose
572 homeostasis,” “increased circulating insulin level,” and “decreased susceptibility to diet-induced
573 obesity”); we observed two significant adiponectin-related gene sets within these meta-gene sets. While
574 these pathway groups had previously been identified in the GWAS DEPICT analysis (**Figure 4**), many of the
575 individual gene sets within these meta-gene sets were not significant in the previous GWAS analysis, such
576 as “insulin resistance,” “abnormal white adipose tissue physiology,” and “abnormal fat cell morphology”
577 (**Supplementary Data 4, Figure 4, Supplementary Figure 16a**), but represent similar biological
578 underpinnings implied by the shared meta-gene sets. Despite their overlap with the GWAS results, these
579 analyses highlight novel genes that fall outside known GWAS loci, based on their strong contribution to
580 the significantly enriched gene sets related to adipocyte and insulin biology (e.g. *MLXIPL*, *ACVR1C*, and
581 *ITIH5*) (**Figure 4**).

582 To focus on novel findings, we conducted pathway analyses after excluding variants from previous
583 WHRadjBMI analyses¹⁰ (**Supplemental Note 2**). Seventy-five loci/genes were included in the EC-DEPICT
584 analysis, and we identified 26 significantly enriched gene sets (13 meta-gene sets). Here, all but one gene
585 set, “lipid particle size”, were related to skeletal biology. This result likely reflects an effect on the pelvic
586 skeleton (hip circumference), shared signaling pathways between bone and fat (such as TGF-beta) and
587 shared developmental origin²⁴ (**Supplementary Data 5, Supplementary Figure 16b**). Many of these
588 pathways were previously found to be significant in the GWAS DEPICT analysis; these findings provide a
589 fully independent replication of their biological relevance for WHRadjBMI.

590 We used PASCAL (**Online Methods**) to further distinguish between enrichment based on *coding-*
591 *only* variant associations (this study) and *regulatory-only* variant associations (up to 20 kb upstream of the
592 gene from a previous GIANT study¹⁰). For completeness, we also compared the coding pathways to those
593 that could be identified in the total previous GWAS effort (using both *coding and regulatory* variants) by
594 PASCAL. The analysis revealed 116 significantly enriched coding pathways (FDR<0.05; **Supplementary**
595 **Table 9**). In contrast, a total of 158 gene sets were identified in the coding+regulatory analysis that
596 included data from the previous GIANT waist GWAS study. Forty-two gene sets were enriched in both
597 analyses. Thus, while we observed high concordance in the $-\log_{10}$ (p-values) between ExomeChip and
598 GWAS gene set enrichment (Pearson's r (coding vs regulatory only) = 0.38, $P < 10^{-300}$; Pearson's r (coding vs
599 coding+regulatory) = 0.51, $P < 10^{-300}$), there are gene sets that seem to be enriched *specifically* for variants
600 in coding regions (e.g., decreased susceptibility to diet-induced obesity, abnormal skeletal morphology)
601 or unique to variants in regulatory regions (e.g. transcriptional regulation of white adipocytes)
602 (**Supplementary Figure 17**).

603 The EC-DEPICT and PASCAL results showed a moderate but strongly significant correlation (for EC-
604 DEPICT and the PASCAL max statistic, $r = .277$ with $p = 9.8 \times 10^{-253}$; for EC-DEPICT and the PASCAL sum
605 statistic, $r = .287$ with $p = 5.42 \times 10^{-272}$). Gene sets highlighted by both methods strongly implicated a role
606 for pathways involved in skeletal biology, glucose homeostasis/insulin signaling, and adipocyte biology.
607 Indeed, we are even more confident in the importance of this core overlapping group of pathways due to
608 their discovery by both methods (**Supplementary Figure 18**).

609 **Cross-trait associations**

610 To assess the relevance of our identified variants with cardiometabolic, anthropometric, and
611 reproductive traits, we conducted association lookups from existing ExomeChip studies of 15 traits
612 (**Supplementary Data 6, Supplementary Figure 19**). Indeed, the clinical relevance of central adiposity is
613 likely to be found in the cascade of impacts such variants have on downstream cardiometabolic

614 disease.^{22,25-29} We found that variants in *STAB1* and *PLCB3* display the greatest number of significant cross-
615 trait associations, each associating with seven different traits ($P < 9.8 \times 10^{-4}$, 0.05/51 variants tested). Of
616 note, these two genes cluster together with *RSPO3*, *DNAH10*, *MNS1*, *COBLL1*, *CCDC92*, and *ITIH3*
617 (**Supplementary Data 6, Supplementary Figure 19**). The WHR-increasing alleles in this cluster of variants
618 exhibit a pattern of increased cardiometabolic risk (e.g. increased fasting insulin [FI], two-hour glucose
619 [TwoHGlu], and triglycerides [TG]; and decreased high-density lipoprotein cholesterol [HDL]), but also
620 decreased BMI. This phenomenon, where variants associated with lower BMI are also associated with
621 increased cardiometabolic risk, has been previously reported.³⁰⁻³⁶ A recent Mendelian Randomization
622 (MR) analysis of the relationship between central adiposity (measured as WHRadjBMI) and
623 cardiometabolic risk factors found central adiposity to be causal.⁹ Using 48 WHR-increasing variants
624 reported in the recent GIANT analysis¹⁰ to calculate a polygenic risk score, Emdin *et al.* found that a 1 SD
625 increase in genetic risk of central adiposity was associated with higher total cholesterol, triglyceride levels,
626 fasting insulin and two-hour glucose, and lower HDL – all indicators of cardiometabolic disease, and also
627 associated with a 1 unit decrease in BMI⁹.

628 We conducted a search in the NHGRI-EBI GWAS Catalog^{37,38} to determine if any of our significant
629 ExomeChip variants are in high LD ($R^2 > 0.7$) with variants associated with traits or diseases not covered by
630 our cross trait lookups (**Supplementary Data 7**). We identified several cardiometabolic traits (adiponectin,
631 coronary heart disease *etc.*) and behavioral traits potentially related to obesity (carbohydrate, fat intake
632 *etc.*) with GWAS associations that were not among those included in cross-trait analyses and nearby one
633 or more of our WHRadjBMI- associated coding variants. Additionally, many of our ExomeChip variants are
634 in LD with GWAS variants associated with other behavioral and neurological traits (schizophrenia, bipolar
635 disorder *etc.*), and inflammatory or autoimmune diseases (Crohn's Disease, multiple sclerosis *etc.*)
636 (**Supplementary Data 7**).

637 Given the established correlation between total body fat percentage and WHR ($R= 0.052$ to
638 0.483)³⁹⁻⁴¹, we examined the association of our top exome variants with both total body fat percentage
639 (BF%) and truncal fat percentage (TF%) available in a sub-sample of up to 118,160 participants of UKBB
640 (**Supplementary Tables 10-11**). Seven of the common novel variants were significantly associated
641 ($P<0.001$, 0.05/48 variants examined) with both BF% and TF% in the sexes-combined analysis (*COBLL1*,
642 *UHRF1BP1*, *WSCD2*, *CCDC92*, *IFI30*, *MPV17L2*, *IZUMO1*). Only one of our tag SNPs, rs7607980 in *COBLL1*,
643 is nearby a known total body fat percentage BF% GWAS locus (rs6738627; $R^2=0.1989$, distance=6751 bp,
644 with our tag SNP)⁴². Two additional variants, rs62266958 in *EFCAB12* and rs224331 in *GDF5*, were
645 significantly associated with TF% in the women-only analysis. Of the nine SNPs associated with at least
646 one of these two traits, all variants displayed much greater magnitude of effect on TF% compared to BF%
647 (**Supplementary Figure 20**).

648 Previous studies have demonstrated the importance of examining common and rare variants
649 within genes with mutations known to cause monogenic diseases^{43,44}. We assessed enrichment of our
650 WHRadjBMI within genes that cause monogenic forms of lipodystrophy) and/or insulin resistance
651 (**Supplementary Data 8**). No significant enrichment was observed (**Supplementary Figure 21**). For
652 lipodystrophy, the lack of significant findings may be due in part to the small number of implicated genes
653 and the relatively small number of variants in monogenic disease-causing genes, reflecting their
654 intolerance of variation.

655 **Genetic architecture of WHRadjBMI coding variants**

656 We used summary statistics from our stage 1 results to estimate the phenotypic variance
657 explained by ExomeChip coding variants. We calculated the variance explained by subsets of SNPs across
658 various significance thresholds ($P< 2\times 10^{-7}$ to 0.2) and conservatively estimated using only independent tag
659 SNPs (**Supplementary Table 12, Online Methods, and Supplementary Figure 22**). The 22 independent
660 significant coding SNPs in stage 1 account for 0.28% of phenotypic variance in WHRadjBMI. For

661 independent variants that reached suggestive significance in stage 1 ($P < 2 \times 10^{-6}$), 33 SNPs explain 0.38% of
662 the variation; however, the 1,786 independent SNPs with a liberal threshold of $P < 0.02$ explain 13 times
663 more variation (5.12%). While these large effect estimates may be subject to winner's curse, for array-
664 wide significant variants, we detected a consistent relationship between effect magnitude and MAF in our
665 stage 2 analyses in UK Biobank and deCODE (**Supplementary Data 1-3**). Notably, the Exomechip coding
666 variants explained less of the phenotypic variance than in our previous GIANT investigation, wherein 49
667 significant SNPs explained 1.4% of the variance in WHRadjBMI. When considering all coding variants on
668 the ExomeChip in men and women together, 46 SNPs with a $P < 2 \times 10^{-6}$ and 5,917 SNPs with a $P < 0.02$ explain
669 0.51% and 13.75% of the variance in WHRadjBMI, respectively. As expected given the design of the
670 ExomeChip, the majority of the variance explained is attributable to rare and low frequency coding
671 variants (independent SNPs with $MAF < 1\%$ and $MAF < 5\%$ explain 5.18% and 5.58%, respectively). However,
672 for rare and low frequency variants, those that passed significance in stage 1 explain only 0.10% of the
673 variance in WHRadjBMI. As in **Figure 2**, these results also indicate that there are additional coding variants
674 associated with WHRadjBMI that remain to be discovered, particularly rare and low frequency variants
675 with larger effects than common variants. Due to observed differences in association strength between
676 women and men, we estimated variance explained for the same set of SNPs in women and men
677 separately. As observed in previous studies¹⁰, there was significantly ($P_{RsqDiff} < 0.002 = 0.05/21$, Bonferroni-
678 corrected threshold) more variance explained in women compared to men at each significance threshold
679 considered (differences ranged from 0.24% to 0.91%).

680 To better understand the potential clinical impact of WHRadjBMI associated variants, we
681 conducted penetrance analysis using the UKBB population (both sexes combined, and men- and women-
682 only). We compared the number of carriers and non-carriers of the minor allele for each of our significant
683 variants in centrally obese and non-obese individuals to determine if there is a significant accumulation
684 of the minor allele in either the centrally obese or non-obese groups (**Online Methods**). Three rare and

685 low frequency variants (MAF \leq 1%) with larger effect sizes (effect size $>$ 0.90) were included in the
686 penetrance analysis using World Health Organization (WHO- obese women WHR $>$ 0.85 and obese men
687 WHR $>$ 0.90) WHR cut-offs for central obesity. Of these, one SNV (rs55920843-ACVR1C; $P_{\text{sex-combined}}=9.25 \times 10^{-5}$;
688 $P_{\text{women}}=4.85 \times 10^{-5}$) showed a statistically significant difference in the number of carriers and non-carriers
689 of the minor allele when the two strata were compared (sex-combined obese carriers=2.2%; non-obese
690 carriers=2.6%; women obese carriers=2.1%; non-obese women carriers=2.6% (**Supplementary Table 13,**
691 **Supplementary Figure 23**). These differences were significant in women, but not in men ($P_{\text{men}} < 5.5 \times 10^{-3}$
692 after Bonferroni correction for 9 tests) and agree with our overall meta-analysis results, where the minor
693 allele (G) was significantly associated with lower WHRadjBMI in women only (**Tables 1 and 2**).

694 **Evidence for functional role of significant variants**

695 ***Drosophila* Knockdown**

696 Considering the genetic evidence of adipose and insulin biology in determining body fat
697 distribution¹⁰, and the lipid signature of the variants described here, we examined whole-body
698 triglycerides levels in adult *Drosophila*, a model organism in which the fat body is an organ functionally
699 analogous to mammalian liver and adipose tissue and triglycerides are the major source of fat storage⁴⁵.
700 Of the 51 genes harboring our 56 significantly associated variants, we identified 27 with *Drosophila*
701 orthologues for functional follow-up analyses. In order to prioritize genes for follow-up, we selected genes
702 with large changes in triglyceride storage levels ($>$ 20% increase or $>$ 40% decrease, as chance alone is
703 unlikely to cause changes of this magnitude, although some decrease is expected) after considering each
704 corresponding orthologue in an existing large-scale screen for adipose with \leq 2 replicates per knockdown
705 strain.⁴⁵ Two orthologues, for *PLXND1* and *DNAH10*, from two separate loci met these criteria. For these
706 two genes, we conducted additional knockdown experiments with \geq 5 replicates using tissue-specific
707 drivers (fat body [cg-Gal4] and neuronal [elav-Gal4] specific RNAi-knockdowns) (**Supplementary Table**
708 **14**). A significant ($P < 0.025$, 0.05/2 orthologues) increase in the total body triglyceride levels was observed

709 in *DNAH10* orthologue knockdown strains for both the fat body and neuronal drivers. However, only the
710 neuronal driver knockdown for *PLXND1* produced a significant change in triglyceride storage. *DNAH10*
711 and *PLXND1* both lie within previous GWAS identified regions. Adjacent genes have been highlighted as
712 likely candidates for the *DNAH10* association region, including *CCDC92* and *ZNF664* based on eQTL
713 evidence. However, our fly knockdown results support *DNAH10* as the causal genes underlying this
714 association. Of note, rs11057353 in *DNAH10* showed suggestive significance after conditioning on the
715 known GWAS variants in nearby *CCDC92* (sex-combined $P_{\text{conditional}}=7.56 \times 10^{-7}$; women-only rs11057353
716 $P_{\text{conditional}}= 5.86 \times 10^{-7}$, **Supplementary Table 6**; thus providing some evidence of multiple causal
717 variants/genes underlying this association signal. Further analyses are needed to determine whether the
718 implicated coding variants from the current analysis are the putatively functional variants, specifically how
719 these variants affect transcription in and around these loci, and exactly how those effects alter biology of
720 relevant human metabolic tissues.

721 ***eQTL Lookups***

722 To gain a better understanding of the potential functionality of novel and low frequency variants,
723 we examined the *cis*-association of the identified variants with expression level of nearby genes in
724 subcutaneous adipose tissue, visceral omental adipose tissue, skeletal muscle and pancreas from GTEx⁴⁶,
725 and assessed whether the exome and eQTL associations implicated the same signal (**Online Methods**,
726 **Supplementary Data 9, Supplementary Table 15**). The lead exome variant was associated with expression
727 level of the coding gene itself for *DAGLB*, *MLXIPL*, *CCDC92*, *MAPKBP1*, *LRRC36* and *UQCC1*. However, at
728 three of these loci (*MLXIPL*, *MAPKBP1*, and *LRRC36*), the lead exome variant is also associated with
729 expression level of additional nearby genes, and at three additional loci, the lead exome variant is only
730 associated with expression level of nearby genes (*HEMK1* at *C3orf18*; *NT5DC2*, *SMIM4* and *TMEM110* at
731 *STAB1/ITIH3*; and *C6orf106* at *UHRF1BP1*). Although detected with a missense variant, these loci are also

732 consistent with a regulatory mechanism of effect as they are significantly associated with expression levels
733 of genes, and the association signal may well be due to LD with nearby regulatory variants.

734 Some of the coding genes implicated by eQTL analyses are known to be involved in adipocyte
735 differentiation or insulin sensitivity: e. g. for *MLXIPL*, the encoded carbohydrate responsive element
736 binding protein is a transcription factor, regulating glucose-mediated induction of *de novo* lipogenesis in
737 adipose tissue, and expression of its *beta*-isoform in adipose tissue is positively correlated with adipose
738 insulin sensitivity^{47,48}. For *CCDC92*, the reduced adipocyte lipid accumulation upon knockdown confirmed
739 the involvement of its encoded protein in adipose differentiation⁴⁹.

740 ***Biological Curation***

741 To gain further insight into the possible functional role of the identified variants, we conducted
742 thorough searches of the literature and publicly available bioinformatics databases (**Supplementary Data**
743 **10-11, Box 1, Online Methods**). Many of our novel low frequency variants are in genes that are intolerant
744 of nonsynonymous mutations (e.g. *ACVR1C*, *DARS2*, *FGFR2*; ExAC Constraint Scores >0.5). Like previously
745 identified GWAS variants, several of our novel coding variants lie within genes that are involved in glucose
746 homeostasis (e.g. *ACVR1C*, *UGGT2*, *ANGPTL4*), angiogenesis (*RASIP1*), adipogenesis (*RAPGEF3*), and lipid
747 biology (*ANGPTL4*, *DAGLB*) (**Supplementary Data 10, Box 1**).

748

749 **DISCUSSION**

750 Our two-staged approach to analysis of coding variants from ExomeChip data in up to 476,546
751 individuals identified a total of 56 array-wide significant variants in 41 independent association signals,
752 including 24 newly identified (23 novel and one independent of known GWAS signals) that influence
753 WHRadjBMI. Nine of these variants were low frequency or rare, indicating an important role for low
754 frequency variants in the polygenic architecture of fat distribution and providing further insights into its

755 underlying etiology. While, due to their rarity, these coding variants only explain a small proportion of the
756 trait variance at a population level, they may, given their predicted role, be more functionally tractable
757 than non-coding variants and have a critical impact at the individual and clinical level. For instance, the
758 association between a low frequency variant (rs11209026; R381Q; MAF<5% in ExAC) located in the *IL23R*
759 gene and multiple inflammatory diseases (such as psoriasis⁵⁰, rheumatoid arthritis⁵¹, ankylosing
760 spondylitis⁵², and inflammatory bowel diseases⁵³) led to the development of new therapies, targeting *IL23*
761 and *IL12* in the same pathway (reviewed in ⁵⁴⁻⁵⁶). Thus, we are encouraged that our associated low
762 frequency coding variants displayed large effect sizes; all but one of the nine novel low frequency variants
763 had an effect size larger than the 49 SNPs reported in Shungin *et al.* 2015, and some of these effect sizes
764 were up to 7-fold larger than those previously reported for GWAS. This finding mirrors results for other
765 cardiometabolic traits⁵⁷, and suggests variants of possible clinical significance with even larger effect and
766 lower frequency variants will likely be detected through larger additional genome-wide scans of many
767 more individuals.

768 We continue to observe sexual dimorphism in the genetic architecture of WHRadjBMI¹¹. Overall,
769 we identified 19 coding variants that display significant sex differences, of which 16 (84%) display larger
770 effects in women compared to men. Of the variants outside of GWAS loci, we reported three (two with
771 MAF<5%) that show a significantly stronger effect in women and two (one with MAF<5%) that show a
772 stronger effect in men. Additionally, genetic variants continue to explain a higher proportion of the
773 phenotypic variation in body fat distribution in women compared to men^{10,11}. Of the novel female (*DSTYK*
774 and *ANGPTL4*) and male (*UGGT2* and *MMP14*) specific signals, only *ANGPTL4* implicated fat distribution
775 related biology associated with both lipid biology and cardiovascular traits (**Box 1**). Sexual dimorphism in
776 fat distribution is apparent from childhood and throughout adult life⁵⁸⁻⁶⁰, and at sexually dimorphic loci,
777 hormones with different levels in men and women may interact with genomic and epigenomic factors to
778 regulate gene activity, though this remains to be experimentally documented. Dissecting the underlying

779 molecular mechanisms of the sexual dimorphism in body fat distribution, and also how it is correlated
780 with – and causing – important comorbidities like T2D and cardiovascular diseases will be crucial for
781 improved understanding of disease risk and pathogenesis.

782 Overall, we observe fewer significant associations between WHRadjBMI and coding variants on
783 the ExomeChip than Turcot *et al.*²⁵ examining the association of low frequency and rare coding variants
784 with BMI. In line with these observations, we identify fewer pathways and cross-trait associations. One
785 reason for fewer WHRadjBMI implicated variants and pathways may be smaller sample size ($N_{\text{WHRadjBMI}} =$
786 $476,546$, $N_{\text{BMI}} = 718,639$), and thus, lower statistical power. Power, however, is likely not the only
787 contributing factor. For example, Turcot *et al.*²⁵ have comparative sample sizes between BMI and that of
788 Marouli *et al.*²² studying height ($N_{\text{height}} = 711,428$). However, greater than seven times the number of
789 coding variants are identified for height than for BMI, indicating that perhaps a number of other factors,
790 including trait architecture, heritability (possibly overestimated in some phenotypes), and phenotype
791 precision, likely all contribute to our study's capacity to identify low frequency and rare variants with large
792 effects. Further, it is possible that the comparative lack of significant findings for WHRadjBMI and BMI
793 compared to height may be a result of higher selective pressure against genetic predisposition to
794 cardiometabolic phenotypes, such as BMI and WHR. As evolutionary theory predicts that harmful alleles
795 will be low frequency⁶¹, we may need larger sample sizes to detect rare variants that have so far escaped
796 selective pressures. Lastly, the ExomeChip is limited by the variants that are present on the chip, which
797 was largely dictated by sequencing studies in European-ancestry populations and a MAF detection criteria
798 of $\sim 0.012\%$. It is likely that through an increased sample size, use of chips designed to detect variation
799 across a range of continental ancestries, high quality, deep imputation with large reference samples (e.g.
800 HRC), and/or alternative study designs, future studies will detect additional variation from the entire allele
801 frequency spectrum that contributes to fat distribution phenotypes.

802 The collected genetic and epidemiologic evidence has now demonstrated that fat distribution (as
803 measured by increased WHRadjBMI) is correlated with increased risk of T2D and CVD, and that this
804 association is likely causal with potential mediation through blood pressure, triglyceride-rich lipoproteins,
805 glucose, and insulin⁹. This observation yields an immediate follow-up question: Which mechanisms
806 regulate depot-specific fat accumulation and are risks for disease, driven by increased visceral or
807 decreased subcutaneous adipose tissue mass (or both)? Pathway analysis identified several novel
808 pathways and gene sets related to metabolism and adipose regulation, bone growth and development
809 we also observed a possible role for adiponectin, a hormone which has been linked to “healthy” expansion
810 of adipose tissue and insulin sensitivity⁶². Similarly, expression/eQTL results support the function and
811 relevance of adipogenesis, adipocyte biology, and insulin signaling, supporting our previous findings for
812 WHRadjBMI¹⁰. We also provide evidence suggesting known biological functions and pathways
813 contributing to body fat distribution (e.g., diet-induced obesity, angiogenesis, bone growth and
814 morphology, and enhanced lipolysis).

815 The ultimate aim of genetic investigations of obesity-related traits, like those presented here, is
816 to identify genomic pathways that are dysregulated leading to obesity pathogenesis, and may result in a
817 myriad of downstream illnesses. Thus, our findings may enhance the understanding of central obesity and
818 identify new molecular targets to avert its negative health consequences. Significant cross-trait
819 associations and additional associations observed in the GWAS Catalog are consistent with expected
820 direction of effect for several traits, i.e. the WHR-increasing allele is associated with higher values of TG,
821 DBP, fasting insulin, TC, LDL and T2D across many significant variants. However, it is worth noting that
822 there are some exceptions. For example, rs9469913-A in *UHRF1BP1* is associated with both increased
823 WHRadjBMI and increased HDL. Also, we identified two variants in *MLXIPL* (rs3812316 and rs35332062),
824 a well-known lipids-associated locus, in which the WHRadjBMI-increasing allele also increases all lipid
825 levels, risk for hypertriglyceridemia, SBP and DBP. However, our findings show a significant and negative

826 association with HbA1C, and nominally significant and negative associations with two-hour glucose,
827 fasting glucose, and Type 2 diabetes, and potential negative associations with biomarkers for liver disease
828 (e.g. gamma glutamyl transpeptidase). Other notable exceptions include *ITIH3* (negatively associated with
829 BMI, HbA1C, LDL and SBP), *DAGLB* (positively associated with HDL), and *STAB1* (negatively associated with
830 TC, LDL, and SBP in cross-trait associations). Therefore, caution in selecting pathways for therapeutic
831 targets is warranted; one must look beyond the effects on central adiposity, but also at the potential
832 cascading effects of related diseases.

833 A seminal finding from this study is the importance of lipid metabolism for body fat distribution.
834 In fact, pathway analyses that highlight enhanced lipolysis, cross-trait associations with circulating lipid
835 levels, existing biological evidence from the literature, and knockdown experiments in *Drosophila*
836 examining triglyceride storage point to novel candidate genes (*ANGPTL4*, *ACVR1C*, *DAGLB*, *MGA*, *RASIP1*,
837 and *IZUMO1*) and new candidates in known regions (*DNAH10*¹⁰ and *MLXIPL*¹⁴) related to lipid biology and
838 its role in fat storage. Newly implicated genes of interest include *ACVR1C*, *MLXIPL*, and *ANGPTL4*, all of
839 which are involved in lipid homeostasis; all are excellent candidate genes for central adiposity. Carriers of
840 inactivating mutations in *ANGPTL4* (*Angiopoietin Like 4*), for example, display low triglyceride levels and
841 low risk of coronary artery disease⁶³. *ACVR1C* encodes the activin receptor-like kinase 7 protein (ALK7), a
842 receptor for the transcription factor TGFB-1, well known for its central role in growth and development in
843 general⁶⁴⁻⁶⁸, and adipocyte development in particular⁶⁸. *ACVR1C* exhibits the highest expression in adipose
844 tissue, but is also highly expressed in the brain⁶⁹⁻⁷¹. In mice, decreased activity of *ACVR1C* upregulates
845 PPAR γ and C/EBP α pathways and increases lipolysis in adipocytes, thus decreasing weight and diabetes in
846 mice^{69,72,73}. Such activity is suggestive of a role for ALK7 in adipose tissue signaling and therefore for
847 therapeutic targets for human obesity. *MLXIPL*, also important for lipid metabolism and postnatal cellular
848 growth, is a transcription factor which activates triglyceride synthesis genes in a glucose-dependent
849 manner^{74,75}. The lead exome variant in this gene is highly conserved, most likely damaging, and is

850 associated with reduced *MLXIPL* expression in adipose tissue. Furthermore, in a recent longitudinal, *in*
851 *vitro* transcriptome analysis of adipogenesis in human adipose-derived stromal cells, gene expression of
852 *MLXIPL* was up-regulated during the maturation of adipocytes, suggesting a critical role in the regulation
853 of adipocyte size and accumulation⁷⁶. However, given our observations on cross-trait associations with
854 variants in *MLXIPL* and diabetes-related traits, development of therapeutic targets must be approached
855 cautiously.

856 Taken together, our 24 novel variants for WHRadjBMI offer new biology, highlighting the
857 importance of lipid metabolism in the genetic underpinnings of body fat distribution. We continue to
858 demonstrate the critical role of adipocyte biology and insulin resistance for central obesity and offer
859 support for potentially causal genes underlying previously identified fat distribution GWAS loci. Notably,
860 our findings offer potential new therapeutic targets for intervention in the risks associated with abdominal
861 fat accumulation, and represents a major advance in our understanding of the underlying biology and
862 genetic architecture of central adiposity.

863
864

865 **ACKNOWLEDGEMENTS**

866 A full list of acknowledgements is provided in the **Supplementary Table 17**. Co-author Yucheng Jia recently
867 passed away while this work was in process. This study was completed as part of the Genetic Investigation
868 of ANthropometric Traits (GIANT) Consortium. This research has been conducted using the UK Biobank
869 resource. Funding for this project was provided by Aase and Ejner Danielsens Foundation, Academy of
870 Finland (102318; 123885; 117844; 40758; 211497; 118590; 139635; 129293; 286284; 134309; 126925;
871 121584; 124282; 129378; 117787; 41071; 137544; 272741), Action on Hearing Loss (G51), ALK-Abelló A/S
872 (Hørsholm-Denmark), American Heart Association (13EIA14220013; 13GRNT16490017;

873 13POST16500011), American Recovery and Reinvestment Act of 2009 (ARRA) Supplement (EY014684-
874 03S1; -04S1; 5RC2HL102419), Amgen, André and France Desmarais Montreal Heart Institute (MHI)
875 Foundation, AstraZeneca, Augustinus Foundation, Australian Government and Government of Western
876 Australia, Australian Research Council Future Fellowship, Becket Foundation, Benzon Foundation, Bernard
877 Wolfe Health Neuroscience Endowment, British Heart Foundation (CH/03/001; RG/14/5/30893;
878 RG/200004; SP/04/002; SP/09/002), BiomarCaRE (278913), Bundesministerium für Bildung und
879 Forschung (Federal Ministry of Education and Research-Germany; German Center for Diabetes Research
880 (DZD); 01ER1206; 01ER1507; 01ER1206; 01ER1507; FKZ: 01EO1501 (AD2-060E); 01ZZ9603; 01ZZ0103;
881 01ZZ0403; 03IS2061A; 03Z1CN22; FKZ 01GI1128), Boehringer Ingelheim Foundation, Boston University
882 School of Medicine, Canada Research Chair program, Canadian Cancer Society Research Institute,
883 Canadian Institutes of Health Research (MOP-82893), Cancer Research UK (C864/A14136; A490/A10124;
884 C8197/A16565), Cebu Longitudinal Health and Nutrition Survey (CLHNS) pilot funds (RR020649;
885 ES010126; DK056350), Center for Non-Communicable Diseases (Pakistan), Central Society for Clinical
886 Research, Centre National de Génotypage (Paris-France), CHDI Foundation (Princeton-USA), Chief
887 Scientist Office of the Scottish Government Health Directorate (CZD/16/6), City of Kuopio and Social
888 Insurance Institution of Finland (4/26/2010), Clarendon Scholarship, Commission of the European
889 Communities; Directorate C-Public Health (2004310), Copenhagen County, County Council of Dalarna,
890 Curtin University of Technology, Dalarna University, Danish Centre for Evaluation and Health Technology
891 Assessment, Danish Council for Independent Research, Danish Diabetes Academy, Danish Heart
892 Foundation, Danish Medical Research Council-Danish Agency for Science Technology and Innovation,
893 Danish Medical Research Council, Danish Pharmaceutical Association, Danish Research Council for
894 Independent Research, Dekker scholarship (2014T001), Dentistry and Health Sciences, Department of
895 Internal Medicine at the University of Michigan, Diabetes Care System West-Friesland, Diabetes Heart
896 Study (R01 HL6734; R01 HL092301; R01 NS058700), Doris Duke Charitable Foundation Clinical Scientist

897 Development Award (2014105), Doris Duke Medical Foundation, Dr. Robert Pflieger Stiftung, Dutch Cancer
898 Society (NKI2009-4363), Dutch Government (NWO 184.021.00; NWO/MaGW VIDI-016-065-318; NWO
899 VICI 453-14-0057; NWO 184.021.007), Dutch Science Organization (ZonMW-VENI Grant 916.14.023),
900 Edith Cowan University, Education and Sports Research Grant (216-1080315-0302); Croatian Science
901 Foundation (grant 8875), Else Kröner-Frsenius-Stiftung (2012_A147), Emil Aaltonen Foundation, Erasmus
902 Medical Center, Erasmus University (Rotterdam), European Research Council Advanced Principal
903 Investigator Award, European Research Council (310644; 268834; 323195; SZ-245 50371-
904 GLUCOSEGENES-FP7-IDEAS-ERC; 293574), Estonian Research Council (IUT20-60), European Union
905 Framework Programme 6 (LSHM_CT_2006_037197; Bloodomics Integrated Project; LSHM-CT-2004-
906 005272; LSHG-CT-2006-018947), European Union Framework Programme 7 (HEALTH-F2-2013-601456;
907 HEALTH-F2-2012-279233; 279153; HEALTH-F3-2010-242244; EpiMigrant; 279143; 313010; 305280;
908 HZ2020 633589; 313010; HEALTH-F2-2011-278913; HEALTH-F4-2007- 201413), European Commission
909 (DG XII), European Community (SOC 98200769 05 F02), European Regional Development Fund to the
910 Centre of Excellence in Genomics and Translational Medicine (GenTransMed), European Union (QLG1-CT-
911 2001-01252; SOC 95201408 05 F02), EVO funding of the Kuopio University Hospital from Ministry of
912 Health and Social Affairs (5254), Eye Birth Defects Foundation Inc., Federal Ministry of Science-Germany
913 (01 EA 9401), Finland's Slottery Machine Association, Finnish Academy (255935; 269517), Finnish
914 Cardiovascular Research Foundation, Finnish Cultural Foundation, Finnish Diabetes Association, Finnish
915 Diabetes Research Foundation, Finnish Foundation for Cardiovascular Research, Finnish Funding Agency
916 for Technology and Innovation (40058/07), Finnish Heart Association, Finnish National Public Health
917 Institute, Fondation Leducq (14CVD01), Food Standards Agency (UK), Framingham Heart Study of the
918 National Heart Lung and Blood Institute of the National Institutes of Health (HHSN268201500001; N02-
919 HL-6-4278), FUSION Study (DK093757; DK072193; DK062370; ZIA-HG000024), General Clinical Research
920 Centre of the Wake Forest School of Medicine (M01 RR07122; F32 HL085989), Genetic Laboratory of the

921 Department of Internal Medicine-Erasmus MC (the Netherlands Genomics Initiative), Genetics and
922 Epidemiology of Colorectal Cancer Consortium (NCI CA137088), German Cancer Aid (70-2488-Ha I),
923 German Diabetes Association, German Research Foundation (CRC 1052 C01; B01; B03), Health and
924 Retirement Study (R03 AG046398), Health Insurance Foundation (2010 B 131), Health Ministry of
925 Lombardia Region (Italy), Helmholtz Zentrum München – German Research Center for Environmental
926 Health, Helse Vest, Home Office (780-TETRA), Hospital Districts of Pirkanmaa; Southern Ostrobothnia;
927 North Ostrobothnia; Central Finland and Northern Savo, Ib Henriksen Foundation, Imperial College
928 Biomedical Research Centre, Imperial College Healthcare NHS Trust, Institute of Cancer Research and The
929 Everyman Campaign, Interuniversity Cardiology Institute of the Netherlands (09.001), Intramural
930 Research Program of the National Institute on Aging, Italian Ministry of Health (GR-2011-02349604), Johns
931 Hopkins University School of Medicine (HHSN268200900041C), Juho Vainio Foundation, Kaiser
932 Foundation Research Institute (HHSN268201300029C), KfH Stiftung Präventivmedizin e.V., KG Jebsen
933 Foundation, Knut and Alice Wallenberg Foundation (Wallenberg Academy Fellow), Knut och Alice
934 Wallenberg Foundation (2013.0126), Kuopio Tampere and Turku University Hospital Medical Funds
935 (X51001), Kuopio University Hospital, Leenaards Foundation, Leiden University Medical Center, Li Ka Shing
936 Foundation (CML), Ludwig-Maximilians-Universität, Lund University, Lundbeck Foundation, Major Project
937 of the Ministry of Science and Technology of China (2017YFC0909700), Marianne and Marcus Wallenberg
938 Foundation, Max Planck Society, Medical Research Council-UK (G0601966; G0700931; G0000934;
939 MR/L01632X/1; MC_UU_12015/1; MC_PC_13048; G9521010D; G1000143; MC_UU_12013/1-9;
940 MC_UU_12015/1; MC_PC_13046; MC_U106179471; G0800270, MR/L01341X/1), MEKOS Laboratories
941 (Denmark), Merck & Co Inc., MESA Family (R01-HL-071205; R01-HL-071051; R01-HL-071250; R01-HL-
942 071251; R01-HL-071252; R01-HL-071258; R01-HL-071259; UL1-RR-025005), Ministry for Health Welfare
943 and Sports (the Netherlands), Ministry of Cultural Affairs (Germany), Ministry of Education and Culture of
944 Finland (627;2004-2011), Ministry of Education Culture and Science (the Netherlands), Ministry of Science

945 and Technology (Taiwan) (MOST 104-2314-B-075A-006 -MY3), Ministry of Social Affairs and Health in
946 Finland, Montreal Heart Institute Foundation, MRC-PHE Centre for Environment and Health, Multi-Ethnic
947 Study of Atherosclerosis (MESA) (N01-HC-95159; N01-HC-95160; N01-HC-95161; N01-HC-95162; N01-HC-
948 95163; N01-HC-95164; N01-HC-95165; N01-HC-95166; N01-HC-95167; N01-HC-95168; N01-HC-95169),
949 Munich Center of Health Sciences (MC-Health), Municipality of Rotterdam (the Netherlands) Murdoch
950 University, National Basic Research Program of China (973 Program 2012CB524900), National Cancer
951 Institute (CA047988; UM1CA182913), National Cancer Research Institute UK, National Cancer Research
952 Network UK, National Center for Advancing Translational Sciences (UL1TR001881), National Center for
953 Research Resources (UL1-TR-000040 and UL1-RR-025005), National Eye Institute of the National Institutes
954 of Health (EY014684, EY-017337), National Health and Medical Research Council of Australia (403981;
955 1021105; 572613), National Heart Lung and Blood Institute (HHSN268200800007C;
956 HHSN268201100037C; HHSN268201200036C; HHSN268201300025C; HHSN268201300026C;
957 HHSN268201300046C; HHSN268201300047C; HHSN268201300048C; HHSN268201300049C;
958 HHSN268201300050C; HHSN268201500001I; HHSN268201700001I; HHSN268201700002I;
959 HHSN268201700003I; HHSN268201700004I; HHSN268201700005I; HL043851; HL080295; HL080467;
960 HL085251; HL087652; HL094535; HL103612; HL105756; HL109946; HL119443; ; HL120393; HL054464;
961 HL054457; HL054481; HL087660; HL086694; HL060944; HL061019; HL060919; HL060944; HL061019;
962 N01HC25195; N01HC55222; N01HC85079; N01HC85080; N01HC85081; N01HC85082; N01HC85083;
963 N01HC85086; N02-HL-6-4278; R21 HL121422-02; R21 HL121422-02; R01 DK089256-05), National Human
964 Genome Research Institute (HG007112), National Institute for Health Research BioResource Clinical
965 Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust
966 and King's College London, National Institute for Health Research Comprehensive Biomedical Research
967 Centre Imperial College Healthcare NHS Trust, National Institute for Health Research (NIHR) (RP-PG-0407-
968 10371), National Institute of Diabetes and Digestive and Kidney Disease (DK063491; DK097524;

969 DK085175; DK087914; 1R01DK8925601; 1R01DK106236-01A1), National Institute of Health Research
970 Senior Investigator, National Institute on Aging (AG023629; NIA U01AG009740; RC2 AG036495; RC4
971 AG039029), National Institute on Minority Health and Health Disparities, National Institutes of Health
972 (NIH) (1R01HG008983-01; 1R21DA040177-01; 1R01HL092577; R01HL128914; K24HL105780;
973 K01HL116770; U01 HL072515-06; U01 HL84756; U01HL105198; U01 GM074518; R01 DK089256-05;
974 R01DK075787; R25 CA94880; P30 CA008748; DK078150; TW005596; HL085144; TW008288; R01-
975 HL093029; U01- HG004729; R01-DK089256; 1R01DK101855-01; K99HL130580; T32-GM067553; U01-
976 DK105561; R01-HL-117078; R01-DK-089256; UO1HG008657; UO1HG06375; UO1AG006781; DK064265;
977 R01DK106621-01; K23HL114724; NS33335; HL57818; R01-DK089256; 2R01HD057194; UO1HG007416;
978 R01DK101855, R01DK075787, T32 GM096911-05; K01 DK107836; R01DK075787; UO1 AG 06781; U01-
979 HG005152, 1F31HG009850-01), National Institute of Neurological Disorders and Stroke, National Key R&D
980 Plan of China (2016YFC1304903), Key Project of the Chinese Academy of Sciences (ZDBS-SSW-DQC-02,
981 ZDRW-ZS-2016-8-1, KJZD-EW-L14-2-2), National Natural Science Foundation of China (81471013;
982 30930081; 81170734; 81321062; 81471013; 81700700), National NIHR Bioresource, National Science
983 Council (Taiwan) (NSC 102-2314-B-075A-002), Netherlands CardioVascular Research Initiative
984 (CVON2011-19), Netherlands Heart Foundation, Netherlands Organisation for Health Research and
985 Development (ZonMW) (113102006), Netherlands Organisation for Scientific Research (NWO)-sponsored
986 Netherlands Consortium for Healthy Aging (050-060-810), Netherlands Organization for Scientific
987 Research (184021007), NHMRC Practitioner Fellowship (APP1103329), NIH through the American
988 Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419), NIHR Biomedical Research Centre at
989 The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, NIHR Cambridge
990 Biomedical Research Centre, NIHR Cambridge Biomedical Research Centre, NIHR Health Protection
991 Research Unit on Health Impact of Environmental Hazards (HPRU-2012-10141), NIHR Leicester
992 Cardiovascular Biomedical Research Unit, NIHR Official Development Assistance (ODA, award 16/136/68),

993 NIHR Oxford Biomedical Research Centre, the European Union FP7 (EpiMigrant, 279143) and H2020
994 programs (iHealth-T2D; 643774), NIHR Senior Investigator, Nordic Centre of Excellence on Systems Biology
995 in Controlled Dietary Interventions and Cohort Studies (SYSDIET) (070014), Northwestern University
996 (HHSN268201300027C), Norwegian Diabetes Association, Novartis, Novo Nordisk Foundation, Nuffield
997 Department of Clinical Medicine Award, Orchid Cancer Appeal, Oxford Biomedical Research Centre, Paavo
998 Nurmi Foundation, Päivikki and Sakari Sohlberg Foundation, Pawsey Supercomputing Centre (funded by
999 Australian Government and Government of Western Australia), Peninsula Research Bank-NIHR Exeter
1000 Clinical Research Facility, Pfizer, Prostate Cancer Research Foundation, Prostate Research Campaign UK
1001 (now Prostate Action), Public Health England, QIMR Berghofer, Raine Medical Research Foundation,
1002 Regione FVG (L.26.2008), Republic of Croatia Ministry of Science, Research Centre for Prevention and
1003 Health-the Capital Region of Denmark, Research Council of Norway, Research Institute for Diseases in the
1004 Elderly (RIDE), Research into Ageing, Robert Dawson Evans Endowment of the Department of Medicine
1005 at Boston University School of Medicine and Boston Medical Center, Science Live/Science Center NEMO,
1006 Scottish Funding Council (HR03006), Sigrid Juselius Foundation, Social Insurance Institution of Finland,
1007 Singapore Ministry of Health's National Medical Research Council (NMRC/STaR/0028/2017), Social
1008 Ministry of the Federal State of Mecklenburg-West Pomerania, State of Bavaria-Germany, State of
1009 Washington Life Sciences Discovery Award (265508) to the Northwest Institute of Genetic Medicine,
1010 Stroke Association, Swedish Diabetes Foundation (2013-024), Swedish Heart-Lung Foundation (20120197;
1011 20120197; 20140422), Swedish Research Council (2012-1397), Swedish Research Council Strategic
1012 Research Network Epidemiology for Health, Swiss National Science Foundation (31003A-143914),
1013 SystemsX.ch (51RTPO_151019), Taichung Veterans General Hospital (Taiwan) (TCVGH-1047319D; TCVGH-
1014 1047311C), Tampere Tuberculosis Foundation, TEKES Grants (70103/06; 40058/07), The Telethon Kids
1015 Institute, Timber Merchant Vilhelm Bangs Foundation, UCL Hospitals NIHR Biomedical Research Centre,
1016 UK Department of Health, Université de Montréal Beaulieu-Saucier Chair in Pharmacogenomics,

1017 University Hospital Regensburg, University of Bergen, University of Cambridge, University of Michigan
1018 Biological Sciences Scholars Program, University of Michigan Internal Medicine Department Division of
1019 Gastroenterology, University of Minnesota (HHSN268201300028C), University of Notre Dame (Australia),
1020 University of Queensland, University of Western Australia (UWA), Uppsala Multidisciplinary Center for
1021 Advanced Computational Science (b2011036), Uppsala University, US Department of Health and Human
1022 Services (HHSN268201100046C; HHSN268201100001C; HHSN268201100002C; HHSN268201100003C;
1023 HHSN268201100004C; HHSN271201100004C), UWA Faculty of Medicine, Velux Foundation, Wellcome
1024 Trust (083948/B/07/Z; 084723/Z/08/Z; 090532; 098381; 098497/Z/12/Z; WT098051; 068545/Z/02;
1025 WT064890; WT086596; WT098017; WT090532; WT098051; WT098017; WT098381; WT098395; 083948;
1026 085475), Western Australian DNA Bank (National Health and Medical Research Council of Australia
1027 National Enabling Facility), Women and Infant's Research Foundation, Yrjö Jahnsson Foundation (56358).

1028 **AUTHORSHIP CONTRIBUTIONS**

1029 Writing Group: LAC, RSF, TMF, MG, HMH, JNH, AEJ, TK, ZK, CML, RJFL, YL, KEN, VT, KLY; Data preparation
1030 group: TA, IBB, TE, SF, MG, HMH, AEJ, TK, DJL, KSL, AEL, RJFL, YL, EM, NGDM, MCMG, PM, MCYN, MAR,
1031 SS, CS, KS, VT, SV, SMW, TWW, KLY, XZ; WHR meta-analyses: PLA, HMH, AEJ, TK, MG, CML, RJFL, KEN, VT,
1032 KLY; Pleiotropy working group: GA, MB, JPC, PD, FD, JCF, HMH, SK, HK, HMH, AEJ, CML, DJL, RJFL, AM, EM,
1033 GM, MIM, PBM, GMP, JRBP, KSR, XS, SW, JW, CJW; Phenome-wide association studies: LB, JCD, TLE, AG,
1034 AM, MIM; Gene-set enrichment analyses: SB, RSF, JNH, ZK, DL, THP; eQTL analyses: CKR, YL, KLM;
1035 Monogenic and syndromic gene enrichment analyses: HMH, AKM; Fly Obesity Screen: AL, JAP; Overseeing
1036 of contributing studies: (1958 Birth Cohort) PD; (Airwave) PE; (AMC PAS) GKH; (Amish) JRO'C; (ARIC) EB;
1037 (ARIC, Add Health) KEN; (BRAVE) EDA, RC; (BRIGHT) PBM; (CARDIA) MF, PJS; (Cebu Longitudinal Health
1038 and Nutrition Survey) KLM; (CHD Exome + Consortium) ASB, JMMH, DFR, JD; (CHES) RV; (Clear/eMERGE
1039 (Seattle)) GPJ; (CROATIA_Korcula) VV, OP, IR; (deCODE) KS, UT; (DHS) DWB; (DIACORE) CAB; (DPS) JT, JL,
1040 MU; (DRSEXTRA) TAL, RR; (EFSOCH) ATH, TMF; (EGCUT) TE; (eMERGE (Seattle)) EBL; (EPIC-Potsdam) MBS,

1041 HB; (EpiHealth) EI, PWF; (EXTEND) ATH, TMF; (Family Heart Study) IBB; (Fenland, EPIC) RAS; (Fenland,
1042 EPIC, InterAct) NJW, CL; (FINRISK) SM; (FINRISK 2007 (T2D)) PJ, VS; (Framingham Heart Study) LAC;
1043 (FUSION) MB, FSC; (FVG) PG; (Generation Scotland) CH, BHS; (Genetic Epidemiology Network of
1044 Arteriopathy (GENOA)) SLRK; (GRAPHIC) NJS; (GSK-STABILITY) DMW, LW, HDW; (Health) AL; (HELIC
1045 MANOLIS) EZ, GD; (HELIC Pomak) EZ, GD; (HUNT-MI) KH, CJW; (Inter99) TH, TJ; (IRASFS) LEW, EKS; (Jackson
1046 Heart Study (JHS)) JGW; (KORA S4) KS, IMH; (Leipzig-Adults) MB, PK; (LOLIPOP-Exome) JCC, JSK; (LOLIPOP-
1047 OmniEE) JCC, JSK; (MESA) JIR, XG; (METSIM) JK, ML; (MONICA-Brianza) GC; (Montreal Heart Institute
1048 Biobank (MHIBB)) MPD, GL, SdD, JCT; (MORGAM Central Laboratory) MP; (MORGAM Data Centre) KK;
1049 (OBB) FK; (PCOS) APM, CML; (PIVUS) CML, LL; (PRIME - Belfast) FK; (PRIME - Lille) PA; (PRIME - Strasbourg)
1050 MM; (PRIME - Toulouse) JF; (PROMIS) DS; (QC) MAR; (RISC) BB, EF, MW; (Rotterdam Study I) AGU, MAI;
1051 (SEARCH) AMD; (SHIP/SHIP-Trend) MD; (SIBS) DFE; (SOLID TIMI-52) DMW; (SORBS) APM, MS, AT; (The
1052 Mount Sinai BioMe Biobank) EPB, RJFL; (The NEO Study) DOMK; (The NHAPC study, The GBTDS study) XL;
1053 (The Western Australian Pregnancy Cohort (Raine) Study) CEP, SM; (TwinsUK) TDS; (ULSAM) APM; (Vejle
1054 Biobank) IB, CC, OP; (WGHS) DIC, PMR; (Women's Health Initiative) PLA; (WTCCC-UKT2D) MIM, KRO; (YFS
1055 TL, OTRa; Genotyping of contributing studies: (1958 Birth Cohort) KES; (Airwave) EE, MPST; (AMC PAS) SS;
1056 (Amish) LMYA, JAP; (ARIC) EWD, MG; (BBMRI-NL) SHV, LB, CMvD, PIWdB; (BRAVE) EDA; (Cambridge
1057 Cancer Studies) JGD; (CARDIA) MF; (CHD Exome + Consortium) ASB, JMMH, DFR, JD, RY(Clear/eMERGE
1058 (Seattle)) GPJ; (CROATIA_Korcula) VV; (DIACORE) CAB, MG; (DPS) AUJ, JL; (DRSEXTRA) PK; (EGCUT) TE;
1059 (EPIC-Potsdam) MBS, KM; (EpiHealth) EI, PWF; (Family Heart Study) KDT; (Fenland, EPIC) RAS; (Fenland,
1060 EPIC, InterAct) NJW, CL; (FUSION) NN; (FVG) IG, AM; (Generation Scotland) CH; (Genetic Epidemiology
1061 Network of Arteriopathy (GENOA)) SLRK, JAS; (GRAPHIC) NJS; (GSK-STABILITY) DMW; (Health) JBJ; (HELIC
1062 MANOLIS) LS; (HELIC Pomak) LS; (Inter99) TH, NG; (KORA) MMN; (KORA S4) KS, HG; (Leipzig-Adults) AM;
1063 (LOLIPOP-Exome) JCC, JSK; (LOLIPOP-OmniEE) JCC, JSK; (MESA) JIR, YDIC, KDT; (METSIM) JK, ML; (Montreal
1064 Heart Institute Biobank (MHIBB)) MPD; (OBB) FK; (PCOS) APM; (PIVUS) CML; (Rotterdam Study I) AGU,

1065 CMG, FR; (SDC) JMJ, HV; (SEARCH) AIMD; (SOLID TIMI-52) DMW; (SORBS) APM; (The Mount Sinai BioMe
1066 Biobank) EPB, RJFL, YL, CS; (The NEO Study) RLG; (The NHAPC study, The GBTDS study) XL, HL, YH; (The
1067 Western Australian Pregnancy Cohort (Raine) Study) CEP, SM; (TUDR) ZA; (TwinsUK) APM; (ULSAM) APM;
1068 (WGHS) DIC, AYC; (Women's Health Initiative) APR; (WTCCC-UKT2D) MIM; (YFS) TL, LPL; Phenotyping of
1069 contributing studies: (Airwave) EE; (AMC PAS) SS; (Amish) LM YA; (ARIC) EWD; (ARIC, Add Health) KEN;
1070 (BBMRI-NL) SHV; (BRAVE) EDA; (BRIGHT) MJC; (CARL) AR, GG; (Cebu Longitudinal Health and Nutrition
1071 Survey) NRL; (CHES) RV, MT; (Clear/eMERGE (Seattle)) GPJ, AAB; (CROATIA_Korcula) OP, IR; (DIACORE)
1072 CAB, BKK; (DPS) AUJ, JL; (EFSOCH) ATH; (EGCUT) EM; (EPIC-Potsdam) HB; (EpiHealth) EI; (EXTEND) ATH;
1073 (Family Heart Study) MFF; (Fenland, EPIC, InterAct) NJW; (FIN-D2D 2007) LM, MV; (FINRISK) SM; (FINRISK
1074 2007 (T2D)) PJ, HS; (Framingham Heart Study) CSF; (Generation Scotland) CH, BHS; (Genetic Epidemiology
1075 Network of Arteriopathy (GENOA)) SLRK, JAS; (GRAPHIC) NJS; (GSK-STABILITY) LW, HDW; (Health) AL, BHT;
1076 (HELIC MANOLIS) LS, AEF, ET; (HELIC Pomak) LS, AEF, MK; (HUNT-MI) KH, OH; (Inter99) TJ, NG; (IRASFS)
1077 LEW, BK; (KORA) MMN; (LASA (BBMRI-NL)) KMAS; (Leipzig-Adults) MB, PK; (LOLIPOP-Exome) JCC, JSK;
1078 (LOLIPOP-OmniEE) JCC, JSK; (MESA) MA; (Montreal Heart Institute Biobank (MHIBB)) GL, KSL, VT;
1079 (MORGAM Data Centre) KK; (OBB) FK, MN; (PCOS) CML; (PIVUS) LL; (PRIME - Belfast) FK; (PRIME - Lille)
1080 PA; (PRIME - Strasbourg) MM; (PRIME - Toulouse) JF; (RISC) BB, EF; (Rotterdam Study I) MAI, CMGFR, MCZ;
1081 (SHIP/SHIP-Trend) NF; (SORBS) MS, AT; (The Mount Sinai BioMe Biobank) EPB, YL, CS; (The NEO Study)
1082 RdM; (The NHAPC study, The GBTDS study) XL, HL, LS, FW; (The Western Australian Pregnancy Cohort
1083 (Raine) Study) CEP; (TUDR) YJH, WJL; (TwinsUK) TDS, KSS; (ULSAM) VG; (WGHS) DIC, PMR; (Women's
1084 Health Initiative) APR; (WTCCC-UKT2D) MIM, KRO; (YFS) TL, OTR; Data analysis of contributing studies:
1085 (1958 Birth Cohort) KES, IN; (Airwave) EE, MPSL; (AMC PAS) SS; (Amish) JRO'C, LMYA, JAP; (ARIC, Add
1086 Health) KEN, KLY, MG; (BBMRI-NL) LB; (BRAVE) RC, DSA; (BRIGHT) HRW; (Cambridge Cancer Studies) JGD,
1087 AE, DJT; (CARDIA) MF, LAL; (CARL) AR, DV; (Cebu Longitudinal Health and Nutrition Survey) YW; (CHD
1088 Exome + Consortium) ASB, JMMH, DFR, RY, PS; (CHES) YJ; (CROATIA_Korcula) VV; (deCODE) VSt, GT; (DHS)

1089 AJC, PM, MCYN; (DIACORE) CAB, MG; (EFSOCH) HY; (EGCUT) TE, RM; (eMERGE (Seattle)) DSC; (ENDO) TK;
1090 (EPIC) JHZ; (EPIC-Potsdam) KM; (EpiHealth) SG; (EXTEND) HY; (Family Heart Study) MFF; (Fenland) JaL;
1091 (Fenland, EPIC) RAS; (Fenland, InterAct) SMW; (Finrisk Extremes and QC) SV; (Framingham Heart Study)
1092 CTL, NLHC; (FVG) IG; (Generation Scotland) CH, JM; (Genetic Epidemiology Network of Arteriopathy
1093 (GENOA)) LFB; (GIANT-Analyst) AEJ; (GRAPHIC) NJS, NGDM, CPN; (GSK-STABILITY) DMW, AS; (Health) JBJ;
1094 (HELIC MANOLIS) LS; (HELIC Pomak) LS; (HUNT-MI) WZ; (Inter99) NG; (IRASFS) BK; (Jackson Heart Study
1095 (JHS)) LAL, JL; (KORA S4) TWW; (LASA (BBMRI-NL)) KMAS; (Leipzig-Adults) AM; (LOLIPOP-Exome) JCC, JSK,
1096 WZ; (LOLIPOP-OmniEE) JCC, JSK, WZ; (MESA) JIR, XG, JY; (METSIM) XS; (Montreal Heart Institute Biobank
1097 (MHIBB)) JCT, GL, KSL, VT; (OBB) AM; (PCOS) APM, TK; (PIVUS) NR; (PROMIS) AR, WZ; (QC GoT2D/T2D-
1098 GENES (FUSION, METSIM, etc)) AEL; (RISC) HY; (Rotterdam Study I) CMG, FR; (SHIP/SHIP-Trend) AT; (SOLID
1099 TIMI-52) DMW, AS; (SORBS) APM; (The Mount Sinai BioMe Biobank) YL, CS; (The NEO Study) RLG; (The
1100 NHAPC study, The GBTDS study) XL, HL, YH; (The Western Australian Pregnancy Cohort (Raine) Study)
1101 CAW; (UK Biobank) ARW; (ULSAM) APM, AM; (WGHS) DIC, AYC; (Women's Health Initiative) PLA, JH;
1102 (WTCCC-UKT2D) WG; (YFS) LPL.

1103 **COMPETING INTERESTS**

1104 The authors declare the following competing interests: ASB holds interest in AstraZeneca, Biogen,
1105 Bioverativ, Merck, Novartis and Pfizer. ASC and CSF are current employees of Merck.
1106 Authors affiliated with deCODE (VSt, GT, UT and KS) are employed by deCODE Genetics/Amgen, I
1107 nc. HDW has the following financial and non-financial competing interests to declare: Research Grants:
1108 Sanofi Aventis; Eli Lilly; NIH; Omthera Pharmaceuticals, Pfizer, Elsal Inc. AstraZeneca; DalCor and Services;
1109 Lecture fees: Sanofi Aventis; Advisory Boards: Acetelion, Sirtex, CSL Boehringer. JD has received grants from
1110 AstraZeneca, Biogen, Merck, Novartis and Pfizer. LMYA and RAS are employee stock holders of

1111 GlaxoSmithKline. MPD received honoraria and holds minor equity in Dalcor. VS has participated in a
1112 conference trip sponsored by Novo Nordisk.

1113 **METHODS**

1114 **Studies**

1115 Stage 1 consisted of 74 studies (12 case/control studies, 59 population-based studies, and five
1116 family studies) comprising 344,369 adult individuals of the following ancestries: 1) European descent (N=
1117 288,492), 2) African (N= 15,687), 3) South Asian (N= 29,315), 4) East Asian (N=6,800), and 5) Hispanic
1118 (N=4,075). Stage 1 meta-analyses were carried out in each ancestry separately and in the all ancestries
1119 group, for both sex-combined and sex-specific analyses. Follow-up analyses were undertaken in 132,177
1120 individuals of European ancestry from the deCODE anthropometric study and UK Biobank (**Supplementary**
1121 **Tables 1-3**). Conditional analyses were performed in the all ancestries and European descent groups.
1122 Informed consent was obtained for participants by the parent study and protocols approved by each
1123 study's institutional review boards.

1124 **Phenotypes**

1125 For each study, WHR (waist circumference divided by hip circumference) was corrected for age,
1126 BMI, and the genomic principal components (derived from GWAS data, the variants with MAF >1% on the
1127 ExomeChip, and ancestry informative markers available on the ExomeChip), as well as any additional
1128 study-specific covariates (e.g. recruiting center), in a linear regression model. For studies with non-related
1129 individuals, residuals were calculated separately by sex, whereas for family-based studies sex was included
1130 as a covariate in models with both men and women. Additionally, residuals for case/control studies were
1131 calculated separately. Finally, residuals were inverse normal transformed and used as the outcome in
1132 association analyses. Phenotype descriptives by study are shown in **Supplementary Table 3**.

1133 **Genotypes and QC**

1134 The majority of studies followed a standardized protocol and performed genotype calling using
1135 the algorithms indicated in **Supplementary Table 2**, which typically included zCall³. For 10 studies
1136 participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium,
1137 the raw intensity data for the samples from seven genotyping centers were assembled into a single project
1138 for joint calling⁴. Study-specific quality control (QC) measures of the genotyped variants were
1139 implemented before association analysis (**Supplementary Tables 1-2**). Furthermore, to assess the
1140 possibility that any significant associations with rare and low-frequency variants could be due to allele
1141 calling in the smaller studies, we performed a sensitivity meta-analysis including all large studies (>5,000
1142 participants) and compared to all studies. We found very high concordance for effect sizes, suggesting
1143 that smaller studies do not bias our results (**Supplementary Fig. 24**).

1144 **Study-level statistical analyses**

1145 Individual cohorts were analyzed for each ancestry separately, in sex-combined and sex-specific
1146 groups, with either RAREMETALWORKER (<http://genome.sph.umich.edu/wiki/RAREMETALWORKER>) or
1147 RVTESTs (<http://zhanxw.github.io/rvtests/>), to associate inverse normal transformed WHRadjBMI with
1148 genotype accounting for cryptic relatedness (kinship matrix) in a linear mixed model. These software
1149 programs are designed to perform score-statistic based rare-variant association analysis, can
1150 accommodate both unrelated and related individuals, and provide single-variant results and variance-
1151 covariance matrices. The covariance matrix captures linkage disequilibrium (LD) relationships between
1152 markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses^{77,78}. Single-
1153 variant analyses were performed for both additive and recessive models.

1154 **Centralized quality-control**

1155 Individual cohorts identified ancestry population outliers based on 1000 Genome Project phase 1
1156 ancestry reference populations. A centralized quality-control procedure implemented in EasyQC⁷⁹ was

1157 applied to individual cohort association summary statistics to identify cohort-specific problems: (1)
1158 assessment of possible errors in phenotype residual transformation; (2) comparison of allele frequency
1159 alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues,
1160 and (3) examination of quantile-quantile (QQ) plots per study to identify any inflation arising from
1161 population stratification, cryptic relatedness and genotype biases.

1162 **Meta-analyses**

1163 Meta-analyses were carried out in parallel by two different analysts at two sites using
1164 RAREMETAL⁷⁷. During the meta-analyses, we excluded variants if they had call rate <95%, Hardy-Weinberg
1165 equilibrium P-value <1x10⁻⁷, or large allele frequency deviations from reference populations (>0.6 for all
1166 ancestries analyses and >0.3 for ancestry-specific population analyses). We also excluded from
1167 downstream analyses markers not present on the Illumina ExomeChip array 1.0, variants on the Y-
1168 chromosome or the mitochondrial genome, indels, multiallelic variants, and problematic variants based
1169 on the Blat-based sequence alignment analyses. Significance for single-variant analyses was defined at an
1170 array-wide level (P<2x10⁻⁷). For all suggestive significant variants from Stage 1, we tested for significant
1171 sex differences. We calculated P_{sexhet} for each SNP, testing for difference between women-specific and
1172 men-specific beta estimates and standard errors using EasyStrata^{11,80}. Each SNP that reached
1173 P_{sexhet}<0.05/# of variants tested (70 variants brought forward from Stage 1, P_{sexhet}<7.14x10⁻⁴) was
1174 considered significant. Additionally, while each individual study was asked to perform association analyses
1175 stratified by race/ethnicity, and adjust for population stratification, all study-specific summary statistics
1176 were meta-analyzed together for our all ancestry meta-analyses. To investigate potential heterogeneity
1177 across ancestries, we did examine ancestry-specific meta-analysis results for our top 70 variants from
1178 stage 1, and found no evidence of significant across-ancestry heterogeneity observed for any of our top
1179 variants (I² values noted in **Supplementary Data 1-3**).

1180 For the gene-based analyses, we applied two sets of criteria to select variants with a MAF<5%
1181 within each ancestry based on coding variant annotation from five prediction algorithms (PolyPhen2,
1182 HumDiv and HumVar, LRT, MutationTaster, and SIFT)^{80,81}. Our broad gene-based tests included nonsense,
1183 stop-loss, splice site, and missense variants annotated as damaging by at least one algorithm mentioned
1184 above. Our strict gene-based tests included only nonsense, stop-loss, splice site, and missense variants
1185 annotated as damaging by all five algorithms. These analyses were performed using the sequence kernel
1186 association test (SKAT) and variable threshold (VT) methods. Statistical significance for gene-based tests
1187 was set at a Bonferroni-corrected threshold of $P < 2.5 \times 10^{-6}$ (0.05/~20,000 genes). All gene-based tests were
1188 performed in RAREMETAL⁷⁷.

1189 **Genomic inflation**

1190 We observed a marked genomic inflation of the test statistics even after controlling for population
1191 stratification (linear mixed model) arising mainly from common markers; λ_{GC} in the primary meta-analysis
1192 (combined ancestries and combined sexes) was 1.06 and 1.37 for all and only common coding and splice
1193 site markers considered herein, respectively (**Supplementary Figures 3, 7 and 13, Supplementary Table**
1194 **16**). Such inflation is expected for a highly polygenic trait like WHRadjBMI, for studies using a non-random
1195 set of variants across the genome, and is consistent with our very large sample size^{79,82,83}.

1196 **Conditional analyses**

1197 The RAREMETAL R-package⁷⁷ was used to identify independent WHRadjBMI association signals
1198 across all ancestries and European meta-analysis results. RAREMETAL performs conditional analyses by
1199 using covariance matrices to distinguish true signals from the shadows of adjacent significant variants in
1200 LD. First, we identified the lead variants ($P < 2 \times 10^{-7}$) based on a 1Mb window centered on the most
1201 significantly associated variant. We then conditioned on the lead variants in RAREMETAL and kept new

1202 lead signals at $P < 2 \times 10^{-7}$ for conditioning in a second round of analysis. The process was repeated until no
1203 additional signal emerged below the pre-specified P-value threshold ($P < 2 \times 10^{-7}$).

1204 To test if the associations detected were independent of the previously published WHRadjBMI
1205 variants^{10,14,16}, we performed conditional analyses in the stage 1 discovery set if the GWAS variant or its
1206 proxy ($r^2 \geq 0.8$) was present on the ExomeChip using RAREMETAL⁷⁷. All variants identified in our meta-
1207 analysis and the previously published variants were also present in the UK Biobank dataset⁸⁴. This dataset
1208 was used as a replacement dataset if a good proxy was not present on the ExomeChip as well as a
1209 replication dataset for the variants present on the ExomeChip. All conditional analyses in the UK Biobank
1210 dataset were performed using SNPTTEST⁸⁵⁻⁸⁷. The conditional analyses were carried out reciprocally,
1211 conditioning on the ExomeChip variant and then the previously published variant. An association was
1212 considered independent of the previously published association if there was a statistically significant
1213 association detected prior to the conditional analysis ($P < 2 \times 10^{-7}$) with both the exome chip variant and the
1214 previously published variant, and the observed association with both or either of the variants disappeared
1215 upon conditional analysis ($P > 0.05$). A conditional p-value between 9×10^{-6} and 0.05 was considered
1216 inconclusive. However, a conditional p-value $< 9 \times 10^{-6}$ was also considered suggestive.

1217

1218 **Stage 2 meta-analyses**

1219 In our Stage 2, we sought to validate a total of 70 variants from Stage 1 that met $P < 2 \times 10^{-6}$ in two
1220 independent studies, the UK Biobank (Release 1⁸⁴) and Iceland (deCODE), comprising 119,572 and 12,605
1221 individuals, respectively (Supplementary Tables 1-3). The same QC and analytical methodology were used
1222 for these studies. Genotyping, study descriptions and phenotype descriptives are provided in
1223 **Supplementary Tables 1-3**. For the combined analysis of Stage 1 plus 2, we used the inverse-variance
1224 weighted fixed effects meta-analysis method. Significant associations were defined as those nominally

1225 significant ($P < 0.05$) in the Stage 2 study and for the combined meta-analysis (Stage 1 plus Stage 2)
1226 significance was set at $P < 2 \times 10^{-7}$ ($0.05 / \sim 250,000$ variants).

1227 **Pathway enrichment analyses: EC-DEPICT**

1228 We adapted DEPICT, a gene set enrichment analysis method for GWAS data, for use with the
1229 ExomeChip ('EC-DEPICT'); this method is also described in a companion manuscript²². DEPICT's primary
1230 innovation is the use of "reconstituted" gene sets, where many different types of gene sets (e.g. canonical
1231 pathways, protein-protein interaction networks, and mouse phenotypes) were extended through the use
1232 of large-scale microarray data (see Pers et al.²¹ for details). EC-DEPICT computes p-values based on
1233 Swedish ExomeChip data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania (ANDIS), and Scania
1234 Diabetes Registry (SDR) cohorts, $N = 11,899$) and, unlike DEPICT, takes as input only the genes directly
1235 containing the significant (coding) variants rather than all genes within a specified amount of linkage
1236 disequilibrium (see **Supplementary Note 2**).

1237 Two analyses were performed for WHRadjBMI ExomeChip: one with all variants $p < 5 \times 10^{-4}$ (49
1238 significant gene sets in 25 meta-gene sets, $FDR < 0.05$) and one with all variants > 1 Mb from known GWAS
1239 loci¹⁰ (26 significant gene sets in 13 meta-gene sets, $FDR < 0.05$). Affinity propagation clustering⁸⁸ was
1240 used to group highly correlated gene sets into "meta-gene sets"; for each meta-gene set, the member
1241 gene set with the best p-value was used as representative for purposes of visualization (see
1242 Supplementary Note). DEPICT for ExomeChip was written using the Python programming language, and
1243 the code can be found at <https://github.com/RebeccaFine/obesity-ec-depict>.

1244 **Pathway enrichment analyses: PASCAL**

1245 We also applied the PASCAL pathway analysis tool²³ to exome-wide association summary statistics
1246 from Stage 1 for all coding variants. The method derives gene-based scores (both SUM and MAX statistics)
1247 and subsequently tests for over-representation of high gene scores in predefined biological pathways. We

1248 used standard pathway libraries from KEGG, REACTOME and BIOCARTA, and also added dichotomized (Z-
1249 score>3) reconstituted gene sets from DEPICT²¹. To accurately estimate SNP-by-SNP correlations even for
1250 rare variants, we used the UK10K data (TwinsUK⁸⁹ and ALSPAC⁹⁰ studies , N=3781). In order to separate
1251 the contribution of regulatory variants from the coding variants, we also applied PASCAL to association
1252 summary statistics of only regulatory variants (20 kb upstream) and regulatory+coding variants from the
1253 Shungin et al¹⁰ study. In this way, we could comment on what is gained by analyzing coding variants
1254 available on ExomeChip arrays. We performed both MAX and SUM estimations for pathway enrichment.
1255 MAX is more sensitive to genesets driven primarily by a single signal, while SUM is better when there are
1256 multiple variant associations in the same gene.

1257 **Monogenic obesity enrichment analyses**

1258 We compiled two lists consisting of 31 genes with strong evidence that disruption causes
1259 monogenic forms of insulin resistance or diabetes; and 8 genes with evidence that disruption causes
1260 monogenic forms of lipodystrophy. To test for enrichment of association, we conducted simulations by
1261 matching each gene with others based on gene length and number of variants tested, to create a matched
1262 set of genes. We generated 1,000 matched gene sets from our data, and assessed how often the number
1263 of variants exceeding set significance thresholds was greater than in our monogenic obesity gene set.

1264 **Variance explained**

1265 We estimated the phenotypic variance explained by the association signals in Stage 1 all
1266 ancestries analyses for men, women, and combined sexes⁹¹. For each associated region, we pruned
1267 subsets of SNPs within 500 kb, as this threshold was comparable with previous studies, of the SNPs with
1268 the lowest P-value and used varying P value thresholds (ranging from 2×10^{-7} to 0.02) from the combined
1269 sexes results. Additionally, we examined all variants and independent variants across a range of MAF
1270 thresholds. The variance explained by each subset of SNPs in each strata was estimated by summing the

1271 variance explained by the individual top coding variants. For the comparison of variance explained
1272 between men and women, we tested for the significance of the differences assuming that the weighted
1273 sum of chi-squared distributed variables tend to a Gaussian distribution ensured by Lyapunov's central
1274 limit theorem.^{91,92}

1275 **Cross-trait lookups**

1276 To carefully explore the relationship between WHRadjBMI and related cardiometabolic,
1277 anthropometric, and reproductive traits, association results for the 51 WHRadjBMI coding SNPs were
1278 requested from existing or on-going meta-analyses from 7 consortia, including ExomeChip data from
1279 GIANT (BMI, height), Global Lipids Genetics Consortium Results (GLGC) (total cholesterol, triglycerides,
1280 HDL-cholesterol, LDL-cholesterol), International Consortium for Blood Pressure (IBPC)⁹³ (systolic and
1281 diastolic blood pressure), Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC)
1282 (glycemic traits), and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (type 2
1283 diabetes).^{22,25-29} For coronary artery disease, we accessed 1000 Genomes Project-imputed GWAS data
1284 released by CARDIoGRAMplusC4D⁹⁴ and for the ReproGen consortium (age at menarche and menopause)
1285 we used a combination of ExomeChip and 1000 Genomes Project-Imputed GWAS data. Heatmaps were
1286 generated in R v3.3.2 using gplots (<https://CRAN.R-project.org/package=gplots>). We used Euclidean
1287 distance based on p-value and direction of effect and complete linkage clustering for the dendrograms.

1288 **GWAS Catalog Lookups**

1289 In order to determine if significant coding variants were associated with any related
1290 cardiometabolic and anthropometric traits, we also searched the NHGRI-EBI GWAS Catalog for previous
1291 variant-trait associations near our lead SNPs (+/- 500 kb). We used PLINK to calculate LD for variants using
1292 ARIC study European participants. All SNVs within the specified regions with an r^2 value > 0.7 were retained
1293 from NHGRI-EBI GWAS Catalog for further evaluation³⁷. Consistent direction of effect was based on WHR-

1294 increasing allele, LD, and allele frequency. Therefore, when a GWAS Catalog variant was not identical or
1295 in high LD ($r^2 > 0.9$) with the WHR variant, and MAF > 0.45 , we do not comment on direction of effect.

1296 **Body-fat percentage associations**

1297 We performed body fat percent and truncal fat percent look-up of 48 of the 56 identified variants
1298 (tables 1 and 2) that were available in the UK Biobank, Release 1⁸⁴, data (notably some of the rare variants
1299 in table 1 and 2 were not available) to further characterize their effects on WHRadjBMI. Genome-wide
1300 association analyses for body fat percent and truncal fat percent were carried out in the UK Biobank. Prior
1301 to analysis, phenotype data were filtered to exclude pregnant or possibly pregnant women, individuals
1302 with body mass index < 15 , and without genetically confirmed European ancestry, resulting in a sample
1303 size of 120,286. Estimated measures of body fat percent and truncal fat percent were obtained using the
1304 Tanita BC418MA body composition analyzer (Tanita, Tokyo, Japan). Individuals were not required to fast
1305 and did not follow any specific instructions prior to the bioimpedance measurements. SNPTTEST was used
1306 to perform the analyses based on residuals adjusted for age, 15 principle components, assessment center
1307 and the genotyping chip⁸⁵.

1308 **Collider bias**

1309 In order to evaluate SNPs for possible collider bias¹⁸, we used results from a recent association
1310 analysis from GIANT on BMI²⁵. For each significant SNP identified in our additive models, WHRadjBMI
1311 associations were corrected for potential bias due to associations between each variant and BMI (See
1312 **Supplementary Note 1** for additional details). Variants were considered robust against collider bias if they
1313 met Bonferroni-corrected significance following correction ($P_{\text{corrected}} < 9.09 \times 10^{-4}$, 0.05/55 variants
1314 examined).

1315 **Drosophila RNAi knockdown experiments**

1316 For each gene in which coding variants were associated with WHRadjBMI in the final combined
1317 meta-analysis ($P < 2 \times 10^{-7}$), its corresponding *Drosophila* orthologues were identified in the Ensembl
1318 ortholog database (www.ensembl.org), when available. *Drosophila* triglyceride content values were
1319 mined from a publicly available genome-wide fat screen data set⁴⁵ to identify potential genes for follow-
1320 up knockdowns. Estimated values represent fractional changes in triglyceride content in adult male flies.
1321 Data are from male progeny resulting from crosses of male UAS-RNAi flies from the Vienna *Drosophila*
1322 Resource Center (VDRC) and Hsp70-GAL4; Tub-GAL8ts virgin females. Two-to-five-day-old males were
1323 sorted into groups of 20 and subjected to two one-hour wet heatshocks four days apart. On the seventh
1324 day, flies were picked in groups of eight, manually crushed and sonicated, and the lysates heat-inactivated
1325 for 10 min in a thermocycler at 95 °C. Centrifuge-cleared supernatants were then used for triglyceride
1326 (GPO Trinder, Sigma) and protein (Pierce) determination. Triglyceride values from these adult-induced
1327 ubiquitous RNAi knockdown individuals were normalized to those obtained in parallel from non-
1328 heatshocked progeny from the very same crosses. The screen comprised one to three biological replicates.
1329 We followed up each gene with a >0.2 increase or >0.4 decrease in triglyceride content.

1330 Orthologues for two genes were brought forward for follow-up, *DNAH10* and *PLXND1*. For both
1331 genes, we generated adipose tissue (*cg-Gal4*) and neuronal (*elav-Gal4*) specific RNAi-knockdown crosses
1332 to knockdown transcripts in a tissue specific manner, leveraging upstream activation sequence (UAS)-
1333 inducible short-hairpin knockdown lines, available through the VDRC (Vienna *Drosophila* Resource
1334 Center). Specifically, *elav-Gal4*, which drives expression of the RNAi construct in post mitotic neurons
1335 starting at embryonic stages all the way to adulthood, was used. *Cg* drives expression in the fat body and
1336 hemocytes starting at embryonic stage 12, all the way to adulthood. We crossed male UAS-RNAi flies and
1337 *elav-GAL4* or *CG-GAL4* virgin female flies. All fly experiments were carried out at 25°C. Five-to-seven-day-
1338 old males were sorted into groups of 20, weighed and homogenated in PBS with 0.05% Tween with Lysing
1339 Matrix D in a beadshaker. The homogenate was heat-inactivated for 10 min in a thermocycler at 70°C.

1340 10 μ l of the homogenate was subsequently used in a triglyceride assay (Sigma, Serum Triglyceride
1341 Determination Kit) which was carried out in duplicate according to protocol, with one alteration: the
1342 samples were cleared of residual particulate debris by centrifugation before absorbance reading.
1343 Resulting triglyceride values were normalized to fly weight and larval/population density. We used the
1344 non-parametric Kruskal-Wallis test to compare wild type with knockdown lines.

1345 **Expression quantitative trait loci (eQTLs) analysis**

1346 We queried the significant variant (Exome coding SNPs)-gene pairs associated with eGenes across
1347 five metabolically relevant tissues (skeletal muscle, subcutaneous adipose, visceral adipose, liver and
1348 pancreas) with at least 70 samples in the GTEx database⁴⁶. For each tissue, variants were selected based
1349 on the following thresholds: the minor allele was observed in at least 10 samples, and the minor allele
1350 frequency was ≥ 0.01 . eGenes, genes with a significant eQTL, are defined on a false discovery rate (FDR)⁹⁵
1351 threshold of ≤ 0.05 of beta distribution-adjusted empirical p-value from FastQTL. Nominal p-values were
1352 generated for each variant-gene pair by testing the alternative hypothesis that the slope of a linear
1353 regression model between genotype and expression deviates from 0. To identify the list of all significant
1354 variant-gene pairs associated with eGenes, a genome-wide empirical p-value threshold⁶⁴, p_t , was defined
1355 as the empirical p-value of the gene closest to the 0.05 FDR threshold. p_t was then used to calculate a
1356 nominal p-value threshold for each gene based on the beta distribution model (from FastQTL) of the
1357 minimum p-value distribution $f(p_{min})$ obtained from the permutations for the gene. For each gene,
1358 variants with a nominal p-value below the gene-level threshold were considered significant and included
1359 in the final list of variant-gene pairs⁶⁴. For each eGene, we also listed the most significantly associated
1360 variants (eSNP). Only these exome SNPs with $r^2 > 0.8$ with eSNPs were considered for the biological
1361 interpretation (Supplementary eQTL GTEx).

1362 We also performed cis-eQTL analysis in 770 METSIM subcutaneous adipose tissue samples as
1363 described in Civelek, et al.⁹⁶ A false discovery rate (FDR) was calculated using all p-values from the cis-

1364 eQTL detection in the q-value package in R. Variants associated with nearby genes at an FDR less than 1%
1365 were considered to be significant (equivalent p-value $< 2.46 \times 10^{-4}$).

1366 For loci with more than one microarray probeset of the same gene associated with the
1367 exome variant, we selected the probeset that provided the strongest LD r^2 between the exome variant
1368 and the eSNP. In reciprocal conditional analysis, we conditioned on the lead exome variant by
1369 including it as a covariate in the cis-eQTL detection and reporting the p-value of the eSNP and vice
1370 versa. We considered the signals to be coincident if both the lead exome variant and the eSNP were no
1371 longer significant after conditioning on the other and the variants were in high pairwise LD ($r^2 > 0.80$).

1372 For loci that also harbored reported GWAS variants, we performed reciprocal conditional analysis
1373 between the GWAS lead variant and the lead eSNP. For loci with more than one reported GWAS variant,
1374 the GWAS lead variant with the strongest LD r^2 with the lead eSNP was reported.

1375 **Penetrance analysis**

1376 Phenotype and genotype data from the UK Biobank (UKBB) were used for the penetrance analysis.
1377 Three of 16 rare and low frequency variants ($MAF \leq 1\%$) detected in the final Stage 1 plus 2 meta-analysis
1378 were available in the UKBB and had relatively larger effect sizes (>0.90). The phenotype data for these
1379 three variants were stratified with respect to waist-to-hip ratio (WHR) using the World Health
1380 Organization (WHO) guidelines. These guidelines consider women and men with WHR greater than 0.85
1381 and 0.90 as obese, respectively. Genotype and allele counts were obtained for the available variants and
1382 these were used to calculate the number of carriers of the minor allele. The number of carriers for women,
1383 men and all combined was then compared between two strata (obese vs. non-obese) using a χ^2 test. The
1384 significance threshold was determined by using a Bonferroni correction for the number of tests performed
1385 ($0.05/9=5.5 \times 10^{-3}$).

1386 **DATA AVAILABILITY**

1387 Summary statistics of all analyses are available at <https://www.broadinstitute.org/collaboration/giant/>.

1388

Box 1. Genes of biological interest harboring WHR-associated variants

PLXND1- (3:129284818, rs2625973, known locus) The major allele of a common non-synonymous variant in Plexin D1 (L1412V, MAF=26.7%) is associated with increased WHRadjBMI (β (SE)= 0.0156 (0.0024), P-value=9.16x10⁻¹¹). *PLXND1* is a semaphorin class 3 and 4 receptor gene, and therefore, is involved in cell to cell signaling and regulation of growth in development for a number of different cell and tissue types, including those in the cardiovascular system, skeleton, kidneys, and the central nervous system⁹⁷⁻¹⁰¹. Mutations in this gene are associated with Moebius syndrome¹⁰²⁻¹⁰⁵, and persistent truncus arteriosus^{99,106}. *PLXND1* is involved in angiogenesis as part of the SEMA and VEGF signalling pathways¹⁰⁷⁻¹¹⁰. *PLXND1* was implicated in the development of T2D through its interaction with *SEMA3E* in mice. *SEMA3E* and *PLXND1* are upregulated in adipose tissue in response to diet-induced obesity, creating a cascade of adipose inflammation, insulin resistance, and diabetes mellitus¹⁰¹. *PLXND1* is highly expressed in adipose (both subcutaneous and visceral) (GTEx). *PLXND1* is highly intolerant of mutations and therefore highly conserved (**Supplementary Data 10**). Last, our lead variant is predicted as damaging or possibly damaging for all algorithms examined (SIFT, Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

ACVR1C- (2:158412701, rs55920843, novel locus) The major allele of a low frequency non-synonymous variant in activin A receptor type 1C (rs55920843, N150H, MAF=1.1%) is associated with increased WHRadjBMI (β (SE)= 0.0652 (0.0105), P-value= 4.81x10⁻¹⁰). *ACVR1C*, also called Activin receptor-like kinase 7 (*ALK7*), is a type I receptor for TGFB (Transforming Growth Factor, Beta-1), and is integral for the activation of SMAD transcription factors; therefore, *ACVR1C* plays an important role in cellular growth and differentiation⁶⁴⁻⁶⁸, including adipocytes⁶⁸. Mouse *Acvr1c* decreases secretion of insulin and

is involved in lipid storage^{69,72,73,69,72,73,111}. *ACVR1C* exhibits the highest expression in adipose tissue, but is also highly expressed in the brain (GTEx)⁶⁹⁻⁷¹. Expression is associated with body fat, carbohydrate metabolism and lipids in both obese and lean individuals⁷⁰. *ACVR1C* is moderately tolerant of mutations (EXAC Constraint Scores: synonymous= -0.86, nonsynonymous = 1.25, LoF = 0.04, **Supplementary Data 10**). Last, our lead variant is predicted as damaging for two of five algorithms examined (LRT and MutationTaster).

FGFR2– (10:123279643, rs138315382, novel locus) The minor allele of a rare synonymous variant in Fibroblast Growth Factor Receptor 2 (rs138315382, MAF=0.09%) is associated with increased WHRadjBMI (β (SE) = 0.258 (0.049), P-value= 1.38×10^{-07}). The extracellular portion of the FGFR2 protein binds with fibroblast growth factors, influencing mitogenesis and differentiation. Mutations in this gene have been associated with many rare monogenic disorders, including skeletal deformities, craniosynostosis, eye abnormalities, and LADD syndrome, as well as several cancers including breast, lung, and gastric cancer. Methylation of *FGFR2* is associated with high birth weight percentile¹¹². *FGFR2* is tolerant of synonymous mutations, but highly intolerant of missense and loss-of-function mutations (ExAC Constraint scores: synonymous=-0.9, missense=2.74, LoF=1.0, **Supplementary Data 10**). Last, this variant is not predicted to be damaging based on any of the 5 algorithms tested.

ANGPTL4 – (19:8429323, rs116843064, novel locus) The major allele of a nonsynonymous low frequency variant in Angiopoietin Like 4 (rs116843064, E40K, EAF=98.1%) is associated with increased WHRadjBMI (β (SE) = 0.064 (0.011) P-value= 1.20×10^{-09}). *ANGPTL4* encodes a glycosylated, secreted protein containing a C-terminal fibrinogen domain. The encoded protein is induced by peroxisome proliferation activators and functions as a serum hormone that regulates glucose homeostasis, triglyceride metabolism^{113,114}, and insulin sensitivity¹¹⁵. *Angptl4*-deficient mice have

hypotriglyceridemia and increased lipoprotein lipase (LPL) activity, while transgenic mice overexpressing *Angptl4* in the liver have higher plasma triglyceride levels and decreased LPL activity¹¹⁶. The major allele of rs116843064 has been previously associated with increased risk of coronary heart disease and increased TG⁶³. *ANGPTL4* is moderately tolerant of mutations (ExAC constraint scores synonymous=1.18, missense=0.21, LoF=0.0, **Supplementary Data 10**). Last, our lead variant is predicted damaging for four of five algorithms (SIFT, Polyphen 2/HDIV, Polyphen2/HVAR, and MutationTaster).

RREB1 – (6:7211818, rs1334576, novel association signal) The major allele of a common non-synonymous variant in the Ras responsive element binding protein 1 (rs1334576, G195R, EAF=56%) is associated with increased WHRadjBMI (β (SE)=0.017 (0.002), P-value=3.9x10⁻¹⁵). This variant is independent of the previously reported GWAS signal in the *RREB1* region (rs1294410; 6:6738752¹⁰). The protein encoded by this gene is a zinc finger transcription factor that binds to RAS-responsive elements (RREs) of gene promoters. It has been shown that the calcitonin gene promoter contains an RRE and that the encoded protein binds there and increases expression of calcitonin, which may be involved in Ras/Raf-mediated cell differentiation¹¹⁷⁻¹¹⁹. The ras responsive transcription factor *RREB1* is a candidate gene for type 2 diabetes associated end-stage kidney disease¹¹⁸. This variant is highly intolerant to loss of function (ExAC constraint score LoF = 1, **Supplementary Data 10**).

DAGLB – (7:6449496, rs2303361, novel locus) The minor allele of a common non-synonymous variant (rs2303361, Q664R, MAF=22%) in *DAGLB* (Diacylglycerol lipase beta) is associated with increased WHRadjBMI (β (SE)= 0.0136 (0.0025), P-value=6.24x10⁻⁸). *DAGLB* is a diacylglycerol (DAG) lipase that catalyzes the hydrolysis of DAG to 2-arachidonoyl-glycerol, the most abundant endocannabinoid in tissues. In the brain, DAGL activity is required for axonal growth during development and for retrograde synaptic signaling at mature synapses (2-AG)¹²⁰. The *DAGLB* variant, rs702485 (7:6449272, r²= 0.306

and $D'=1$ with rs2303361) has been previously associated with high-density lipoprotein cholesterol (HDL) previously. Pathway analysis indicate a role in the triglyceride lipase activity pathway¹²¹. *DAGLB* is tolerant of synonymous mutations, but intolerant of missense and loss of function mutations (ExAC Constraint scores: synonymous=-0.76, missense=1.07, LoF=0.94, **Supplementary Data 10**). Last, this variant is not predicted to be damaging by any of the algorithms tested.

MLXIPL (7:73012042, rs35332062 and 7:73020337, rs3812316, known locus) The major alleles of two common non-synonymous variants (A358V, MAF=12%; Q241H, MAF=12%) in *MLXIPL* (MLX interacting protein like) are associated with increased WHRadjBMI (β (SE)= 0.02 (0.0033), P-value= 1.78×10^{-9} ; β (SE)= 0.0213 (0.0034), P-value= 1.98×10^{-10}). These variants are in strong linkage disequilibrium ($r^2=1.00$, $D'=1.00$, 1000 Genomes CEU). This gene encodes a basic helix-loop-helix leucine zipper transcription factor of the Myc/Max/Mad superfamily. This protein forms a heterodimeric complex and binds and activates carbohydrate response element (ChoRE) motifs in the promoters of triglyceride synthesis genes in a glucose-dependent manner^{74,75}. This gene is possibly involved in the growth hormone signaling pathway and lipid metabolism. The WHRadjBMI-associated variant rs3812316 in this gene has been associated with the risk of non-alcoholic fatty liver disease and coronary artery disease^{74,122,123}. Furthermore, Williams-Beuren syndrome (an autosomal dominant disorder characterized by short stature, abnormal weight gain, various cardiovascular defects, and mental retardation) is caused by a deletion of about 26 genes from the long arm of chromosome 7 including *MLXIPL*. *MLXIPL* is generally intolerant to variation, and therefore conserved (ExAC Constraint scores: synonymous = 0.48, missense=1.16, LoF=0.68, **Supplementary Data 10**). Last, both variants reported here are predicted as possible or probably damaging by one of the algorithms tested (PolyPhen).

RAPGEF3 (12:48143315, rs145878042, novel locus) The major allele of a low frequency non-synonymous variant in Rap Guanine-Nucleotide-Exchange Factor (GEF) 3 (rs145878042, L300P, MAF=1.1%) is associated with increased WHRadjBMI (β (SE)=0.085 (0.010), P-value = $7.15E^{-17}$). *RAPGEF3* codes for an intracellular cAMP sensor, also known as Epac (the Exchange Protein directly Activated by Cyclic AMP). Among its many known functions, *RAPGEF3* regulates the ATP sensitivity of the KATP channel involved in insulin secretion¹²⁴, may be important in regulating adipocyte differentiation¹²⁵⁻¹²⁷, plays an important role in regulating adiposity and energy balance¹²⁸. *RAPGEF3* is tolerant of mutations (ExAC Constraint Scores: synonymous = -0.47, nonsynonymous = 0.32, LoF = 0, **Supplementary Data 10**). Last, our lead variant is predicted as damaging or possibly damaging for all five algorithms examined (SIFT, Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

TBX15 (1:119427467, rs61730011, known locus) The major allele of a low frequency non-synonymous variant in T-box 15 (rs61730011, M460R, MAF=4.3%) is associated with increased WHRadjBMI (β (SE)=0.041(0.005)). T-box 15 (*TBX15*) is a developmental transcription factor expressed in adipose tissue, but with higher expression in visceral adipose tissue than in subcutaneous adipose tissue, and is strongly downregulated in overweight and obese individuals¹²⁹. *TBX15* negatively controls depot-specific adipocyte differentiation and function¹³⁰ and regulates glycolytic myofiber identity and muscle metabolism¹³¹. *TBX15* is moderately intolerant of mutations and therefore conserved (ExAC Constraint Scores: synonymous = 0.42, nonsynonymous = 0.65, LoF = 0.88, **Supplementary Data 10**). Last, our lead variant is predicted as damaging or possibly damaging for four of five algorithms (Polyphen2/HDIV, Polyphen2/HVAR, LRT, MutationTaster).

1390 **REFERENCES**

- 1391 1. Pischon, T. *et al.* General and abdominal adiposity and risk of death in Europe. *N Engl J Med* **359**,
1392 2105-20 (2008).
- 1393 2. Wang, Y., Rimm, E.B., Stampfer, M.J., Willett, W.C. & Hu, F.B. Comparison of abdominal adiposity
1394 and overall obesity in predicting risk of type 2 diabetes among men. *Am J Clin Nutr* **81**, 555-63
1395 (2005).
- 1396 3. Canoy, D. Distribution of body fat and risk of coronary heart disease in men and women. *Curr Opin*
1397 *Cardiol* **23**, 591-8 (2008).
- 1398 4. Snijder, M.B. *et al.* Associations of hip and thigh circumferences independent of waist
1399 circumference with the incidence of type 2 diabetes: the Hoorn Study. *Am J Clin Nutr* **77**, 1192-7
1400 (2003).
- 1401 5. Yusuf, S. *et al.* Obesity and the risk of myocardial infarction in 27,000 participants from 52
1402 countries: a case-control study. *Lancet* **366**, 1640-9 (2005).
- 1403 6. Mason, C., Craig, C.L. & Katzmarzyk, P.T. Influence of central and extremity circumferences on all-
1404 cause mortality in men and women. *Obesity (Silver Spring)* **16**, 2690-5 (2008).
- 1405 7. Karpe, F. & Pinnick, K.E. Biology of upper-body and lower-body adipose tissue--link to whole-body
1406 phenotypes. *Nat Rev Endocrinol* **11**, 90-100 (2015).
- 1407 8. Manolopoulos, K.N., Karpe, F. & Frayn, K.N. Gluteofemoral body fat as a determinant of metabolic
1408 health. *Int J Obes (Lond)* **34**, 949-59 (2010).
- 1409 9. Emdin, C.A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2
1410 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626-634 (2017).
- 1411 10. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature*
1412 **518**, 187-96 (2015).

- 1413 11. Winkler, T.W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size
1414 and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
- 1415 12. Wen, W. *et al.* Genome-wide association studies in East Asians identify new loci for waist-hip ratio
1416 and waist circumference. *Sci Rep* **6**, 17958 (2016).
- 1417 13. Gao, C. *et al.* A Comprehensive Analysis of Common and Rare Variants to Identify Adiposity Loci
1418 in Hispanic Americans: The IRAS Family Study (IRASFS). *PLoS One* **10**, e0134649 (2015).
- 1419 14. Graff, M. *et al.* Genome-wide physical activity interactions in adiposity - A meta-analysis of
1420 200,452 adults. *PLoS Genet* **13**, e1006528 (2017).
- 1421 15. Justice, A.E. *et al.* Genome-wide meta-analysis of 241,258 adults accounting for smoking
1422 behaviour identifies novel loci for obesity traits. *Nat Commun* **8**, 14977 (2017).
- 1423 16. Ng, M.C.Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of
1424 genome-wide association studies in individuals of African ancestry: African Ancestry
1425 Anthropometry Genetics Consortium. *PLoS Genet* **13**, e1006719 (2017).
- 1426 17. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature*
1427 **518**, 197-206 (2015).
- 1428 18. Aschard, H., Vilhjalmsson, B.J., Joshi, A.D., Price, A.L. & Kraft, P. Adjusting for heritable covariates
1429 can bias effect estimates in genome-wide association studies. *Am J Hum Genet* **96**, 329-39 (2015).
- 1430 19. Day, F.R., Loh, P.R., Scott, R.A., Ong, K.K. & Perry, J.R. A Robust Example of Collider Bias in a
1431 Genetic Association Study. *Am J Hum Genet* **98**, 392-3 (2016).
- 1432 20. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-
1433 analysis for rare variants. *Bioinformatics* **30**, 2828-9 (2014).
- 1434 21. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene
1435 functions. *Nat Commun* **6**, 5890 (2015).

- 1436 22. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**,
1437 186-190 (2017).
- 1438 23. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation
1439 of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput Biol* **12**, e1004714
1440 (2016).
- 1441 24. Kawai, M., de Paula, F.J. & Rosen, C.J. New insights into osteoporosis: the bone-fat connection. *J*
1442 *Intern Med* **272**, 317-29 (2012).
- 1443 25. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that
1444 control energy intake and expenditure in obesity. *Nat Genet* **50**, 26-41 (2018).
- 1445 26. Liu, D.J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. **49**, 1758-
1446 1766 (2017).
- 1447 27. Kraja, A.T. *et al.* New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000
1448 Individuals. *Circ Cardiovasc Genet* **10**(2017).
- 1449 28. Mahajan, A. *et al.* Identification and functional characterization of G6PC2 coding variants
1450 influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet*
1451 **11**, e1004876 (2015).
- 1452 29. Manning, A. *et al.* A Low-Frequency Inactivating AKT2 Variant Enriched in the Finnish Population
1453 Is Associated With Fasting Insulin Levels and Type 2 Diabetes Risk. *Diabetes* **66**, 2019-2032 (2017).
- 1454 30. Zhao, W. *et al.* Identification of new susceptibility loci for type 2 diabetes and shared etiological
1455 pathways with coronary heart disease. **49**, 1450-1457 (2017).
- 1456 31. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture
1457 and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
- 1458 32. Ng, M.C. *et al.* Meta-analysis of genome-wide association studies in African Americans provides
1459 insights into the genetic architecture of type 2 diabetes. *PLoS Genet* **10**, e1004517 (2014).

- 1460 33. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic
1461 architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-44 (2014).
- 1462 34. Saxena, R. *et al.* Genome-wide association study identifies a novel locus contributing to type 2
1463 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* **62**, 1746-55 (2013).
- 1464 35. Cook, J.P. & Morris, A.P. Multi-ethnic genome-wide association study identifies novel locus for
1465 type 2 diabetes susceptibility. *Eur J Hum Genet* **24**, 1175-80 (2016).
- 1466 36. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale
1467 association analysis. *Nat Genet* **42**, 579-89 (2010).
- 1468 37. Burdett, T. *et al.* The NHGRI-EBI Catalog of published genome-wide association studies. v1.0 edn
1469 Vol. 2015 (2015).
- 1470 38. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci
1471 for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
- 1472 39. Lutoslawska, G. *et al.* Relationship between the percentage of body fat and surrogate indices of
1473 fatness in male and female Polish active and sedentary students. *J Physiol Anthropol* **33**, 10 (2014).
- 1474 40. Verma, M., Rajput, M., Sahoo, S.S., Kaur, N. & Rohilla, R. Correlation between the percentage of
1475 body fat and surrogate indices of obesity among adult population in rural block of Haryana. *J*
1476 *Family Med Prim Care* **5**, 154-9 (2016).
- 1477 41. Pereira, P.F. *et al.* [Measurements of location of body fat distribution: an assessment of colinearity
1478 with body mass, adiposity and stature in female adolescents]. *Rev Paul Pediatr* **33**, 63-71 (2015).
- 1479 42. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic
1480 disease risk. *Nat Commun* **7**, 10495 (2016).
- 1481 43. Chambers, J.C. *et al.* Common genetic variation near MC4R is associated with waist circumference
1482 and insulin resistance. *Nat Genet* **40**, 716-8 (2008).

- 1483 44. Nead, K.T. *et al.* Contribution of common non-synonymous variants in PCSK1 to body mass index
1484 variation and risk of obesity: a systematic review and meta-analysis with evidence from up to 331
1485 175 individuals. *Hum Mol Genet* **24**, 3582-94 (2015).
- 1486 45. Pospisilik, J.A. *et al.* Drosophila genome-wide obesity screen reveals hedgehog as a determinant
1487 of brown versus white adipose cell fate. *Cell* **140**, 148-60 (2010).
- 1488 46. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
1489 multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
- 1490 47. Baraille, F., Planchais, J., Dentin, R., Guilmeau, S. & Postic, C. Integration of ChREBP-Mediated
1491 Glucose Sensing into Whole Body Metabolism. *Physiology (Bethesda)* **30**, 428-37 (2015).
- 1492 48. Kursawe, R. *et al.* Decreased transcription of ChREBP-alpha/beta isoforms in abdominal
1493 subcutaneous adipose tissue of obese adolescents with prediabetes or early type 2 diabetes:
1494 associations with insulin resistance and hyperglycemia. *Diabetes* **62**, 837-44 (2013).
- 1495 49. Lotta, L.A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage
1496 capacity in the pathogenesis of human insulin resistance. *Nat Genet* **49**, 17-26 (2017).
- 1497 50. Cargill, M. *et al.* A large-scale genetic association study confirms IL12B and leads to the
1498 identification of IL23R as psoriasis-risk genes. *Am J Hum Genet* **80**, 273-90 (2007).
- 1499 51. Hazlett, J., Stamp, L.K., Merriman, T., Highton, J. & Hessian, P.A. IL-23R rs11209026 polymorphism
1500 modulates IL-17A expression in patients with rheumatoid arthritis. *Genes Immun* **13**, 282-7 (2012).
- 1501 52. Karaderi, T. *et al.* Association between the interleukin 23 receptor and ankylosing spondylitis is
1502 confirmed by a new UK case-control study and meta-analysis of published series. *Rheumatology*
1503 (*Oxford*) **48**, 386-9 (2009).
- 1504 53. Duerr, R.H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel
1505 disease gene. *Science* **314**, 1461-3 (2006).

- 1506 54. Abdollahi, E., Tavasolian, F., Momtazi-Borojeni, A.A., Samadi, M. & Rafatpanah, H. Protective role
1507 of R381Q (rs11209026) polymorphism in IL-23R gene in immune-mediated diseases: A
1508 comprehensive review. *J Immunotoxicol* **13**, 286-300 (2016).
- 1509 55. Abraham, C., Dulai, P.S., Vermeire, S. & Sandborn, W.J. Lessons Learned From Trials Targeting
1510 Cytokine Pathways in Patients With Inflammatory Bowel Diseases. *Gastroenterology* **152**, 374-388
1511 e4 (2017).
- 1512 56. Molinelli, E., Campanati, A., Ganzetti, G. & Offidani, A. Biologic Therapy in Immune Mediated
1513 Inflammatory Disease: Basic Science and Clinical Concepts. *Curr Drug Saf* **11**, 35-43 (2016).
- 1514 57. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-7 (2016).
- 1515 58. Wells, J.C. Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* **21**, 415-
1516 30 (2007).
- 1517 59. Loomba-Albrecht, L.A. & Styne, D.M. Effect of puberty on body composition. *Curr Opin Endocrinol*
1518 *Diabetes Obes* **16**, 10-5 (2009).
- 1519 60. Rogol, A.D., Roemmich, J.N. & Clark, P.A. Growth at puberty. *J Adolesc Health* **31**, 192-200 (2002).
- 1520 61. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-45 (2012).
- 1521 62. Stern, J.H., Rutkowski, J.M. & Scherer, P.E. Adiponectin, Leptin, and Fatty Acids in the
1522 Maintenance of Metabolic Homeostasis through Adipose Tissue Crosstalk. *Cell Metab* **23**, 770-84
1523 (2016).
- 1524 63. Dewey, F.E. *et al.* Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J*
1525 *Med* **374**, 1123-33 (2016).
- 1526 64. Bondestam, J. *et al.* cDNA cloning, expression studies and chromosome mapping of human type I
1527 serine/threonine kinase receptor ALK7 (ACVR1C). *Cytogenet Cell Genet* **95**, 157-62 (2001).

- 1528 65. Jornvall, H., Blokzijl, A., ten Dijke, P. & Ibanez, C.F. The orphan receptor serine/threonine kinase
1529 ALK7 signals arrest of proliferation and morphological differentiation in a neuronal cell line. *J Biol*
1530 *Chem* **276**, 5140-6 (2001).
- 1531 66. Kim, B.C. *et al.* Activin receptor-like kinase-7 induces apoptosis through activation of MAPKs in a
1532 Smad3-dependent mechanism in hepatoma cells. *J Biol Chem* **279**, 28458-65 (2004).
- 1533 67. Watanabe, R. *et al.* The MH1 domains of smad2 and smad3 are involved in the regulation of the
1534 ALK7 signals. *Biochem Biophys Res Commun* **254**, 707-12 (1999).
- 1535 68. Kogame, M. *et al.* ALK7 is a novel marker for adipocyte differentiation. *J Med Invest* **53**, 238-45
1536 (2006).
- 1537 69. Murakami, M. *et al.* Expression of activin receptor-like kinase 7 in adipose tissues. *Biochem Genet*
1538 **51**, 202-10 (2013).
- 1539 70. Carlsson, L.M. *et al.* ALK7 expression is specific for adipose tissue, reduced in obesity and
1540 correlates to factors implicated in metabolic disease. *Biochem Biophys Res Commun* **382**, 309-14
1541 (2009).
- 1542 71. Carithers, L.J. & Moore, H.M. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank*
1543 **13**, 307-8 (2015).
- 1544 72. Yogosawa, S., Mizutani, S., Ogawa, Y. & Izumi, T. Activin receptor-like kinase 7 suppresses lipolysis
1545 to accumulate fat in obesity through downregulation of peroxisome proliferator-activated
1546 receptor gamma and C/EBPalpha. *Diabetes* **62**, 115-23 (2013).
- 1547 73. Yogosawa, S. & Izumi, T. Roles of activin receptor-like kinase 7 signaling and its target, peroxisome
1548 proliferator-activated receptor gamma, in lean and obese adipocytes. *Adipocyte* **2**, 246-50 (2013).
- 1549 74. Seifi, M., Ghasemi, A., Namipashaki, A. & Samadikuchaksaraei, A. Is C771G polymorphism of MLX
1550 interacting protein-like (MLXIPL) gene a novel genetic risk factor for non-alcoholic fatty liver
1551 disease? *Cell Mol Biol (Noisy-le-grand)* **60**, 37-42 (2014).

- 1552 75. Cairo, S., Merla, G., Urbinati, F., Ballabio, A. & Reymond, A. WBSR14, a gene mapping to the
1553 Williams-Beuren syndrome deleted region, is a new member of the Mlx transcription factor
1554 network. *Hum Mol Genet* **10**, 617-27 (2001).
- 1555 76. Ambele, M.A., Dessels, C., Durandt, C. & Pepper, M.S. Genome-wide analysis of gene expression
1556 during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene
1557 expression during adipocyte differentiation. *Stem Cell Res* **16**, 725-34 (2016).
- 1558 77. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet* **46**, 200-4
1559 (2014).
- 1560 78. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population
1561 analysis. *Bioinformatics* **28**, 2543-5 (2012).
- 1562 79. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat*
1563 *Protoc* **9**, 1192-212 (2014).
- 1564 80. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature*
1565 **518**, 187-196 (2015).
- 1566 81. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**,
1567 185-90 (2014).
- 1568 82. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12
1569 (2011).
- 1570 83. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies
1571 additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
- 1572 84. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range
1573 of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
- 1574 85. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for
1575 genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

- 1576 86. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven
1577 common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
- 1578 87. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev*
1579 *Genet* **11**, 499-511 (2010).
- 1580 88. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6
1581 (2007).
- 1582 89. Moayyeri, A., Hammond, C.J., Valdes, A.M. & Spector, T.D. Cohort Profile: TwinsUK and healthy
1583 ageing twin study. *Int J Epidemiol* **42**, 76-85 (2013).
- 1584 90. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal
1585 Study of Parents and Children. *Int J Epidemiol* **42**, 111-27 (2013).
- 1586 91. Kutalik, Z., Whittaker, J., Waterworth, D., Beckmann, J.S. & Bergmann, S. Novel method to
1587 estimate the phenotypic variation explained by genome-wide association studies reveals large
1588 fraction of the missing heritability. *Genet Epidemiol* **35**, 341-9 (2011).
- 1589 92. Billingsley, P. *Probability and measure*, xii, 622 p. (Wiley, New York, 1986).
- 1590 93. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated
1591 with blood pressure and hypertension. *Nat Genet* **48**, 1151-61 (2016).
- 1592 94. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis
1593 of coronary artery disease. *Nat Genet* **47**, 1121-30 (2015).
- 1594 95. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S*
1595 *A* **100**, 9440-5 (2003).
- 1596 96. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am*
1597 *J Hum Genet* **100**, 428-443 (2017).
- 1598 97. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**, D222-6
1599 (2015).

- 1600 98. Toyofuku, T. *et al.* Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses
1601 angiogenesis via Plexin-D1. *EMBO J* **26**, 1373-84 (2007).
- 1602 99. Gitler, A.D., Lu, M.M. & Epstein, J.A. PlexinD1 and semaphorin signaling are required in endothelial
1603 cells for cardiovascular development. *Dev Cell* **7**, 107-16 (2004).
- 1604 100. Luchino, J. *et al.* Semaphorin 3E suppresses tumor cell death triggered by the plexin D1
1605 dependence receptor in metastatic breast cancers. *Cancer Cell* **24**, 673-85 (2013).
- 1606 101. Shimizu, I. *et al.* Semaphorin3E-induced inflammation contributes to insulin resistance in dietary
1607 obesity. *Cell Metab* **18**, 491-504 (2013).
- 1608 102. Verzijl, H.T., van der Zwaag, B., Cruysberg, J.R. & Padberg, G.W. Mobius syndrome redefined: a
1609 syndrome of rhombencephalic maldevelopment. *Neurology* **61**, 327-33 (2003).
- 1610 103. Verzijl, H.T., van der Zwaag, B., Lammens, M., ten Donkelaar, H.J. & Padberg, G.W. The
1611 neuropathology of hereditary congenital facial palsy vs Mobius syndrome. *Neurology* **64**, 649-53
1612 (2005).
- 1613 104. Fujita, M., Reinhart, F. & Neutra, M. Convergence of apical and basolateral endocytic pathways at
1614 apical late endosomes in absorptive cells of suckling rat ileum in vivo. *J Cell Sci* **97 (Pt 2)**, 385-94
1615 (1990).
- 1616 105. Briegel, W. Neuropsychiatric findings of Mobius sequence -- a review. *Clin Genet* **70**, 91-7 (2006).
- 1617 106. Ta-Shma, A. *et al.* Isolated truncus arteriosus associated with a mutation in the plexin-D1 gene.
1618 *Am J Med Genet A* **161A**, 3115-20 (2013).
- 1619 107. Mazzotta, C. *et al.* Plexin-D1/Semaphorin 3E pathway may contribute to dysregulation of vascular
1620 tone control and defective angiogenesis in systemic sclerosis. *Arthritis Res Ther* **17**, 221 (2015).
- 1621 108. Yang, W.J. *et al.* Semaphorin-3C signals through Neuropilin-1 and PlexinD1 receptors to inhibit
1622 pathological angiogenesis. *EMBO Mol Med* **7**, 1267-84 (2015).

- 1623 109. Zygumnt, T. *et al.* Semaphorin-PlexinD1 signaling limits angiogenic potential via the VEGF decoy
1624 receptor sFlt1. *Dev Cell* **21**, 301-14 (2011).
- 1625 110. Kim, J., Oh, W.J., Gaiano, N., Yoshida, Y. & Gu, C. Semaphorin 3E-Plexin-D1 signaling regulates
1626 VEGF function in developmental angiogenesis via a feedback mechanism. *Genes Dev* **25**, 1399-411
1627 (2011).
- 1628 111. Bertolino, P. *et al.* Activin B receptor ALK7 is a negative regulator of pancreatic beta-cell function.
1629 *Proc Natl Acad Sci U S A* **105**, 7246-51 (2008).
- 1630 112. Haworth, K.E. *et al.* Methylation of the FGFR2 gene is associated with high birth weight centile in
1631 humans. *Epigenomics* **6**, 477-91 (2014).
- 1632 113. Chi, X. *et al.* Angiopoietin-like 4 Modifies the Interactions between Lipoprotein Lipase and Its
1633 Endothelial Cell Transporter GPIHBP1. *J Biol Chem* **290**, 11865-77 (2015).
- 1634 114. Catoire, M. *et al.* Fatty acid-inducible ANGPTL4 governs lipid metabolic response to exercise. *Proc*
1635 *Natl Acad Sci U S A* **111**, E1043-52 (2014).
- 1636 115. van Raalte, D.H. *et al.* Angiopoietin-like protein 4 is differentially regulated by glucocorticoids and
1637 insulin in vitro and in vivo in healthy humans. *Exp Clin Endocrinol Diabetes* **120**, 598-603 (2012).
- 1638 116. Koster, A. *et al.* Transgenic angiopoietin-like (angptl)4 overexpression and targeted disruption of
1639 angptl4 and angptl3: regulation of triglyceride metabolism. *Endocrinology* **146**, 4943-50 (2005).
- 1640 117. Thiagalingam, A. *et al.* RREB-1, a novel zinc finger protein, is involved in the differentiation
1641 response to Ras in human medullary thyroid carcinomas. *Mol Cell Biol* **16**, 5335-45 (1996).
- 1642 118. Bonomo, J.A. *et al.* The ras responsive transcription factor RREB1 is a novel candidate gene for
1643 type 2 diabetes associated end-stage kidney disease. *Hum Mol Genet* **23**, 6441-7 (2014).
- 1644 119. Thiagalingam, A., Lengauer, C., Baylin, S.B. & Nelkin, B.D. RREB1, a ras responsive element binding
1645 protein, maps to human chromosome 6p25. *Genomics* **45**, 630-2 (1997).

- 1646 120. Bisogno, T. *et al.* Cloning of the first sn1-DAG lipases points to the spatial and temporal regulation
1647 of endocannabinoid signaling in the brain. *J Cell Biol* **163**, 463-8 (2003).
- 1648 121. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat*
1649 *Genet* **45**, 1274-83 (2013).
- 1650 122. Kooner, J.S. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma
1651 triglycerides. *Nat Genet* **40**, 149-51 (2008).
- 1652 123. Pan, L.A. *et al.* G771C Polymorphism in the MLXIPL Gene Is Associated with a Risk of Coronary
1653 Artery Disease in the Chinese: A Case-Control Study. *Cardiology* **114**, 174-8 (2009).
- 1654 124. Kang, G., Leech, C.A., Chepurny, O.G., Coetzee, W.A. & Holz, G.G. Role of the cAMP sensor Epac
1655 as a determinant of KATP channel ATP sensitivity in human pancreatic beta-cells and rat INS-1
1656 cells. *J Physiol* **586**, 1307-19 (2008).
- 1657 125. Ji, Z., Mei, F.C. & Cheng, X. Epac, not PKA catalytic subunit, is required for 3T3-L1 preadipocyte
1658 differentiation. *Front Biosci (Elite Ed)* **2**, 392-8 (2010).
- 1659 126. Martini, C.N., Plaza, M.V. & Vila Mdel, C. PKA-dependent and independent cAMP signaling in 3T3-
1660 L1 fibroblasts differentiation. *Mol Cell Endocrinol* **298**, 42-7 (2009).
- 1661 127. Petersen, R.K. *et al.* Cyclic AMP (cAMP)-mediated stimulation of adipocyte differentiation requires
1662 the synergistic action of Epac- and cAMP-dependent protein kinase-dependent processes. *Mol*
1663 *Cell Biol* **28**, 3804-16 (2008).
- 1664 128. Yan, J. *et al.* Enhanced leptin sensitivity, reduced adiposity, and improved glucose homeostasis in
1665 mice lacking exchange protein directly activated by cyclic AMP isoform 1. *Mol Cell Biol* **33**, 918-26
1666 (2013).
- 1667 129. Gesta, S. *et al.* Evidence for a role of developmental genes in the origin of obesity and body fat
1668 distribution. *Proc Natl Acad Sci U S A* **103**, 6676-81 (2006).

- 1669 130. Gesta, S. *et al.* Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and
1670 mitochondrial respiration. *Proc Natl Acad Sci U S A* **108**, 2771-6 (2011).
- 1671 131. Lee, K.Y. *et al.* Tbx15 controls skeletal muscle fibre-type determination and muscle metabolism.
1672 *Nat Commun* **6**, 8054 (2015).
- 1673
- 1674
- 1675

1676 **FIGURES**

1677 **Figure 1. Summary of meta-analysis study design and workflow.** Abbreviations:

1678 EUR- European, AFR- African, SAS- South Asian, EAS- East Asian, and HIS- Hispanic/Latino ancestry.

1679 **Figure 2.** Minor allele frequency compared to estimated effect. This scatter plot displays the relationship
1680 between minor allele frequency (MAF) and the estimated effect (β) for each significant coding variant in
1681 our meta-analyses. All novel WHRadjBMI variants are highlighted in orange, and variants identified only
1682 in models that assume recessive inheritance are denoted by diamonds and only in sex-specific analyses
1683 by triangles. Eighty percent power was calculated based on the total sample size in the Stage 1+2 meta-
1684 analysis and $P=2 \times 10^{-7}$. Estimated effects are shown in original units (cm/cm) calculated by using effect
1685 sizes in standard deviation (SD) units times SD of WHR in the ARIC study (sexes combined=0.067,
1686 men=0.052, women=0.080).

1687 **Figure 3.** Regional association plots for known loci with novel coding signals. Point color reflects r^2
1688 calculated from the ARIC dataset. In a) there are two independent variants in *RSPO3* and *KIAA0408*, as
1689 shown by conditional analysis. In b) we have a variant in *RREB1* that is independent of the GWAS variant
1690 rs1294421.

1691 **Figure 4.** Heat maps showing DEPICT gene set enrichment results. For any given square, the color indicates
1692 how strongly the corresponding gene (shown on the x-axis) is predicted to belong to the reconstituted
1693 gene set (y-axis). This value is based on the gene's z-score for gene set inclusion in DEPICT's reconstituted
1694 gene sets, where red indicates a higher and blue a lower z-score. To visually reduce redundancy and
1695 increase clarity, we chose one representative "meta-gene set" for each group of highly correlated gene
1696 sets based on affinity propagation clustering (**Online Methods, Supplementary Note 2**). Heatmap
1697 intensity and DEPICT P-values (see P-values in **Supplementary Data 4-5**) correspond to the most
1698 significantly enriched gene set within the meta-gene set. Annotations for the genes indicate (1) the minor

1699 allele frequency of the significant ExomeChip (EC) variant (shades of blue; if multiple variants, the lowest-
1700 frequency variant was kept), (2) whether the variant's P-value reached array-wide significance ($<2 \times 10^{-7}$)
1701 or suggestive significance ($<5 \times 10^{-4}$) (shades of purple), (3) whether the variant was novel, overlapping
1702 "relaxed" GWAS signals from Shungin et al.¹⁰ (GWAS $P < 5 \times 10^{-4}$), or overlapping "stringent" GWAS signals
1703 (GWAS $P < 5 \times 10^{-8}$) (shades of pink), and (4) whether the gene was included in the gene set enrichment
1704 analysis or excluded by filters (shades of brown/orange) (Online Methods and Supplementary
1705 Information). Annotations for the gene sets indicate if the meta-gene set was found significant (shades of
1706 green; FDR < 0.01 , < 0.05 , or not significant) in the DEPICT analysis of GWAS results from Shungin et al.

1707

1708

1709 TABLES

1710 **Table 1. Association results for Combined Sexes.** Association results based on an additive or recessive model for coding variants that met array-wide significance ($P < 2 \times 10^{-7}$) in the sex-combined
 1711 meta-analyses.

Locus (+/-1Mb of a given variant)	Chr:Position (GRCh37) ^b	rsID	EA	OA	Gene ^c	Amino Acid Change ^c	If locus is known, nearby (< 1 MB) published variant(s) ^d	N	EAf	β^e	SE	P-value	P-value for Sex- heterogeneity ^f	Other Criteria For Sig ^h
Variants in Novel Loci														
All Ancestry Additive model Sex-combined analyses														
1	2:158412701	rs55920843	T	G	<i>ACVR1C</i>	N150H	-	455,526	0.989	0.065	0.011	4.8E-10	1.7E-07	
2	3:50597092	rs1034405	G	A	<i>C3orf18</i>	A162V	-	455,424	0.135	0.016	0.003	1.9E-07	8.8E-01	G,C
3	4:120528327	rs3733526	G	A	<i>PDE5A</i>	A41V	-	461,521	0.187	0.015	0.003	2.6E-08	5.2E-03	
4	6:26108117	rs146860658	T	C	<i>HIST1H1T</i>	A69T	-	217,995	0.001	0.229	0.042	4.3E-08	6.3E-01	S
5	7:6449496	rs2303361	C	T	<i>DAGLB</i>	Q664R	-	475,748	0.221	0.014	0.003	6.2E-08	3.4E-03	G
6	10:123279643	rs138315382	T	C	<i>FGFR2</i>	synonymous	-	236,962	0.001	0.258	0.049	1.4E-07	1.1E-01	G,S
7	11:65403651	rs7114037	C	A	<i>PCNXL3</i>	H1822Q	-	448,861	0.954	0.029	0.005	1.8E-08	4.4E-01	
8	12:48143315	rs145878042	A	G	<i>RAPGEF3</i>	L300P	-	470,513	0.990	0.085	0.010	7.2E-17	7.3E-03	
9	12:108618630	rs3764002	C	T	<i>WSCD2</i>	T266I	-	474,637	0.737	0.014	0.002	9.8E-10	5.5E-01	
10	15:42032383	rs17677991	G	C	<i>MGA</i>	P1523A	-	469,874	0.345	0.015	0.002	3.5E-11	9.1E-01	
11	16:4432029	rs3810818	A	C	<i>VASN</i>	E384A	-	424,163	0.231	0.016	0.003	2.0E-09	3.3E-01	
	16:4445327	rs3747579	C	T	<i>CORO7</i>	R193Q	-	453,078	0.299	0.018	0.002	2.2E-13	4.3E-02	
	16:4484396	rs1139653	A	T	<i>DNAJA3</i>	N75Y	-	434,331	0.284	0.015	0.002	4.3E-10	1.4E-01	
12	19:49232226	rs2287922	A	G	<i>RASIP1</i>	R601C	-	430,272	0.494	0.014	0.002	1.6E-09	3.7E-02	
	19:49244220	rs2307019	G	A	<i>IZUMO1</i>	A333V	-	476,147	0.558	0.012	0.002	4.7E-08	3.9E-02	
13	20:42965811	rs144098855	T	C	<i>R3HDML</i>	P5L	-	428,768	0.001	0.172	0.032	9.7E-08	1.0E+00	G

European Ancestry Additive model Sex-combined analyses														
14	1:173802608	rs35515638	G	A	<i>DARS2</i>	K196R	-	352,646	0.001	0.201	0.038	1.4E-07	6.0E-02	G
15	14:58838668	rs1051860	A	G	<i>ARID4A</i>	synonymous	-	367,079	0.411	0.013	0.002	2.2E-08	1.3E-01	
16	15:42115747	rs3959569	C	G	<i>MAPKBP1</i>	R1240H	-	253,703	0.349	0.017	0.003	2.0E-08	6.3E-01	

Variants in Previously Identified Loci

All Ancestry Additive model Sex-combined analyses														
1	1:119427467	rs61730011	A	C	<i>TBX15</i>	M566R	rs2645294, rs12731372, rs12143789, rs1106529	441,461	0.957	0.041	0.005	2.2E-14	6.7E-01	
	1:119469188	rs10494217	T	G		H156N		472,259	0.174	0.018	0.003	1.4E-10	6.0E-01	
2	1:154987704	rs141845046	C	T	<i>ZBTB7B</i>	P190S	rs905938	476,440	0.976	0.037	0.007	3.8E-08	7.9E-07	C
3	2:165551201	rs7607980	T	C	<i>COBLL1</i>	N941D	rs1128249, rs10195252, rs12692737, rs12692738, rs17185198	389,883	0.879	0.026	0.004	1.6E-13	3.0E-30	
4	2:188343497	rs7586970	T	C	<i>TFPI</i>	N221S	rs1569135	452,638	0.697	0.016	0.002	3.0E-12	6.3E-01	
5	3:52558008	rs13303	T	C	<i>STAB1</i>	M113T	rs2276824	470,111	0.445	0.019	0.002	5.5E-18	6.7E-02	
	3:52833805	rs3617	C	A	<i>ITIH3</i>	Q315K		452,150	0.541	0.015	0.002	1.6E-12	4.0E-01	C
6	3:129137188	rs62266958	C	T	<i>EFCAB12</i>	R197H	rs10804591	476,382	0.936	0.036	0.004	8.3E-17	9.3E-05	
	3:129284818	rs2625973	A	C	<i>PLXND1</i>	L1412V		476,338	0.733	0.016	0.002	9.2E-11	1.6E-05	
7	4:89625427	rs1804080	G	C	<i>HERC3</i>	E946Q	rs9991328	446,080	0.838	0.021	0.003	1.5E-12	4.1E-06	
	4:89668859	rs7657817	C	T	<i>FAM13A</i>	V443I		476,383	0.815	0.016	0.003	5.0E-09	9.6E-05	
8	5:176516631	rs1966265	A	G	<i>FGFR4</i>	V10I	rs6556301	455,246	0.236	0.023	0.003	1.7E-19	2.1E-01	
9	6:7211818	rs1334576^g	G	A	<i>RREB1</i>	G195R	rs1294410	451,044	0.565	0.017	0.002	3.9E-15	1.5E-01	
10	6:34827085	rs9469913	A	T	<i>UHRF1BP1</i>	Q984H	rs1776897	309,684	0.847	0.021	0.004	1.2E-08	2.7E-01	C
11	6:127476516	rs1892172	A	G	<i>RSPO3</i>	synonymous	rs11961815, rs72959041, rs1936805	476,358	0.543	0.031	0.002	2.6E-47	7.7E-09	
	6:127767954	rs139745911^g	A	G	<i>KIAA0408</i>	P504S		391,469	0.010	0.103	0.012	6.8E-19	2.0E-04	
12	7:73012042	rs35332062	G	A	<i>MLXIPL</i>	A358V	rs6976930	451,158	0.880	0.020	0.003	1.8E-09	1.5E-01	
	7:73020337	rs3812316	C	G		Q241H		454,738	0.881	0.021	0.003	2.0E-10	5.8E-02	

13	10:95931087	rs17417407	T	G	<i>PLCE1</i>	R240L	rs10786152	476,475	0.173	0.018	0.003	2.5E-11	5.9E-01	
14	11:64031241	rs35169799	T	C	<i>PLCB3</i>	S778L	rs11231693	476,457	0.061	0.034	0.004	9.1E-15	1.3E-04	
15	12:123444507	rs58843120	G	T	<i>ABDB9</i>	F92L	rs4765219, rs863750	466,498	0.987	0.053	0.009	1.3E-08	3.5E-01	
	12:124265687	rs11057353	T	C	<i>DNAH10</i>	S228P		476,360	0.373	0.018	0.002	2.1E-16	2.7E-08	
	12:124330311	rs34934281	C	T		T1785M		476,395	0.889	0.025	0.003	2.9E-14	3.1E-08	
	12:124427306	rs11057401	T	A	<i>CCDC92</i>	S53C		467,649	0.695	0.029	0.002	7.3E-37	5.5E-11	
16	15:56756285	rs1715919	G	T	<i>MNS1</i>	Q55P	rs8030605	476,274	0.096	0.023	0.004	8.8E-11	2.7E-02	
17	16:67397580	rs9922085	G	C	<i>LRRC36</i>	R101P	rs6499129	469,474	0.938	0.034	0.005	3.8E-13	5.9E-01	
	16:67409180	rs8052655	G	A		G388S		474,035	0.939	0.034	0.005	5.5E-13	4.0E-01	
18	19:18285944	rs11554159	A	G	<i>IFI30</i>	R76Q	rs12608504	476,389	0.257	0.015	0.002	3.5E-10	3.1E-03	
	19:18304700	rs874628	G	A	<i>MPV17L2</i>	M72V		476,388	0.271	0.015	0.002	1.2E-10	2.5E-03	
19	20:33971914	rs4911494	T	C	<i>UQCC1</i>	R51Q	rs224333	451,064	0.602	0.018	0.002	2.5E-16	1.5E-03	
	20:34022387	rs224331	A	C	<i>GDF5</i>	S276A		345,805	0.644	0.017	0.003	1.8E-11	3.2E-03	

All Ancestry Recessive model Sex-combined analyses

20	17:17425631	rs897453	C	T	<i>PEMT</i>	V58L	rs4646404	476,546	0.569	0.025	0.004	4.1E-11	8.2E-01	
----	-------------	----------	---	---	-------------	------	-----------	---------	-------	-------	-------	----------------	---------	--

European Ancestry Additive model Sex-combined analyses

6	3:129293256	rs2255703	T	C	<i>PLXND1</i>	M870V	rs10804591	420,520	0.620	0.014	0.002	3.1E-09	1.6E-04	
---	-------------	-----------	---	---	---------------	-------	------------	---------	-------	-------	-------	----------------	----------------	--

1712 Abbreviations: GRCh37=human genome assembly build37;rsID=based on dbSNP; VEP=Ensembl Variant Effect Predictor toolset; GTEx=Genotype-Tissue Expression project;SD=standard deviation; SE=standard error;N=sample size;

1713 EAF=effect allele frequency; EA=effect allele; OA=other allele.

1714 a Coding variants refer to variants located in the exons and splicing junction regions.

1715 b Variant positions are reported according to Human assembly build 37 and their alleles are coded based on the positive strand.

1716 c The gene the variant falls in and amino acid change from the most abundant coding transcript is shown (protein annotation is based on VEP toolset and transcript abundance from GTEx database).

1717 d Previously published variants within +/-1Mb are from Shungin et al.¹⁰, except for rs6976930 and rs10786152 from Graff et al.¹⁴ and rs6499129 from Ng. et al.¹⁶.

1718 e Effect size is based on standard deviation (SD) per effect allele

1719 f P-value for sex heterogeneity, testing for difference between women-specific and men-specific beta estimates and standard errors, was calculated using EasyStrata: Winkler, T.W. et al. EasyStrata: evaluation and visualization of

1720 stratified genome-wide association meta-analysis data. Bioinformatics 2015; 31, 259-61.PMID: 25260699. Bolded P-values met significance threshold after bonferonni correction (P-value<7.14E-04; i.e. 0.05/70 variants).

1721 g **rs1334576 in *RREB1*** is a new signal in a known locus that is independent from the known signal, rs1294410; **rs139745911 in *KIAA0408*** is a new signal in a known locus that is independent from all known signals rs11961815, rs72959041,

1722 rs1936805, in a known locus (see Supplementary 8A/B).

1723 h Each flag indicates a that a secondary criteria for significance may not be met, G- P-value $> 5 \times 10^{-8}$ (GWAS significant), C- Association Signal was not robust against collider bias; S- variant was not available in stage 2 studies for validation
1724 of Stage 1 association.
1725

1726 **Table 2. Association results for Sex-stratified analyses.** Association results based on an additive or recessive model for coding variants that met array-wide significance ($P < 2 \times 10^{-7}$) in the sex-
 1727 specific meta-analyses and reach bonferonni corrected P-value for sex heterogeneity ($P_{\text{sexhet}} < 7.14 \times 10^{-4}$).

Locus (+/-1Mb of a given variant)	Chr:Position (GRCh37) ^c	rsID	EA	OA	Gene ^d	Amino Acid Change ^d	In sex-combined analyses ^e	If locus is known, nearby (< 1 MB) published variant(s) ^f	P-value for Sex-heterogeneity ^g	Men					Women					Other Criteria For Sig ^j	
										N	EAf	β^h	SE	P	N	EAf	β^h	SE	P		
Variants in Novel Loci																					
All Ancestry Additive model Men only analyses																					
1	13:96665697	rs148108950	A	G	<i>UGGT2</i>	P175L	No	-	1.5E-06	203,009	0.006	0.130	0.024	6.1E-08	221,390	0.004	-0.044	0.027	1.1E-01	G	
2	14:23312594	rs1042704	A	G	<i>MMP14</i>	D273N	No	-	2.6E-04	226,646	0.202	0.021	0.004	2.6E-08	250,018	0.197	0.002	0.004	6.1E-01		
All Ancestry Additive model Women only analyses																					
3	1:205130413	rs3851294	G	A	<i>DSTYK</i>	C641R	No	-	9.8E-08	225,803	0.914	-0.005	0.005	3.4E-01	249,471	0.912	0.034	0.005	4.5E-11		
4	2:158412701	rs55920843	T	G	<i>ACVR1C</i>	N150H	Yes	-	1.7E-07	210,071	0.989	0.006	0.015	7.2E-01	245,808	0.989	0.113	0.014	1.7E-15		
5	19:8429323	rs116843064	G	A	<i>ANGPTL4</i>	E40K	No	-	1.3E-07	203,098	0.981	-0.017	0.011	1.4E-01	243,351	0.981	0.064	0.011	1.2E-09		
Variants in Previously Identified Loci																					
All Ancestry Additive model Women only analyses																					
1	1:154987704	rs141845046	C	T	<i>ZBTB7B</i>	P190S	Yes	rs905938	7.9E-07	226,709	0.975	0.004	0.010	6.9E-01	250,084	0.977	0.070	0.010	2.3E-13		
2	2:165551201	rs7607980	T	C	<i>COBLL1</i>	N941D	Yes	rs1128249, rs10195252, rs12692737, rs12692738, rs17185198	3.0E-30	173,600	0.880	-0.018	0.005	5.8E-04	216,636	0.878	0.062	0.005	6.7E-39		
3	3:129137188	rs62266958	C	T	<i>EFCAB12</i>	R197H	Yes	rs10804591	9.3E-05	226,690	0.937	0.018	0.006	3.1E-03	250,045	0.936	0.051	0.006	8.1E-18		
	3:129284818	rs2625973	A	C	<i>PLXND1</i>	L1412V	Yes		1.6E-05	226,650	0.736	0.005	0.003	1.9E-01	250,023	0.730	0.025	0.003	8.2E-14		
	3:129293256	rs2255703	T	C		M870V	Yes		5.0E-04	226,681	0.609	0.003	0.003	3.1E-01	250,069	0.602	0.018	0.003	1.9E-09		
4	4:89625427	rs1804080	G	C	<i>HERC3</i>	E946Q	Yes	rs9991328	4.1E-06	222,556	0.839	0.008	0.004	6.6E-02	223,877	0.837	0.034	0.004	2.1E-16		

	4:89668859	rs7657817	C	T	<i>FAM13A</i>	V443I	Yes		9.6E-05	226,680	0.816	0.006	0.004	1.5E-01	242,970	0.815	0.026	0.004	5.9E-12
5	6:127476516	rs1892172	A	G	<i>RSPO3</i>	synonymous	Yes	rs11961815, rs72959041, rs1936805	7.7E-09	226,677	0.541	0.018	0.003	5.6E-10	250,034	0.545	0.042	0.003	3.4E-48
	6:127767954	rs139745911	A	G	<i>KIAA0408</i>	P504S	Yes		2.0E-04	188,079	0.010	0.057	0.017	6.8E-04	205,203	0.010	0.143	0.016	5.9E-19
6	11:64031241	rs35169799	T	C	<i>PLCB3</i>	S778L	Yes	rs11231693	1.3E-04	226,713	0.061	0.016	0.006	9.6E-03	250,097	0.061	0.049	0.006	6.7E-16
7	12:124265687	rs11057353	T	C	<i>DNAH10</i>	S228P	Yes	rs4765219, rs863750	2.7E-08	226,659	0.370	0.005	0.003	8.3E-02	250,054	0.376	0.029	0.003	3.1E-22
	12:124330311	rs34934281	C	T		T1785M	Yes		3.1E-08	226,682	0.891	0.006	0.005	1.9E-01	250,066	0.887	0.043	0.005	1.4E-20
	12:124427306	rs11057401	T	A	<i>CCDC92</i>	S53C	Yes		5.5E-11	223,324	0.701	0.013	0.003	4.3E-05	244,678	0.689	0.043	0.003	1.0E-41

1728 Abbreviations: GRCh37=human genome assembly build 37;rsID=based on dbSNP; VEP=Ensembl Variant Effect Predictor toolset; GTEx=Genotype-Tissue Expression project; SD=standard deviation; SE=standard error;N=sample size; EA=effect
1729 allele; OA=other allele; EAF=effect allele frequency.

1730 a Coding variants refer to variants located in the exons and splicing junction regions.

1731 b Bonferonni corrected Pvalue for the number of SNPs tested for sex-heterogeneity is <7.14E-04 i.e. 0.05/70 variants.

1732 c Variant positions are reported according to Human assembly build 37 and their alleles are coded based on the positive strand.

1733 d The gene the variant falls in and amino acid change from the most abundant coding transcript is shown (protein annotation is based on VEP toolset and transcript abundance from GTEx database).

1734 e Variant was also identified as array-wide significant in the sex-combined analyses.

1735 f Previously published variants within +/-1Mb are from Shungin D et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015; 518, 187–196 doi:10.1038/nature14132 (PMID 25673412).

1736 g P-value for sex heterogeneity, testing for difference between women-specific and men-specific beta estimates and standard errors, was calculated using EasyStrata: Winkler, T.W. et al. EasyStrata: evaluation and visualization of stratified
1737 genome-wide association meta-analysis data. Bioinformatics 2015: 31, 259-61. PMID: 25260699.

1738 h Effect size is based on standard deviation (SD) per effect allele

1739 i rs139745911 in KIAA0408 is a new signal in a known locus that is independent from all known signals rs11961815, rs72959041, rs1936805, in a known locus (see Supplementary 8A/B).

1740 j Each flag indicates a that a secondary criteria for significance may not be met, G- P-value > 5x10-8 (GWAS significant), C- Association Signal was not robust against collider bias; S- variant was not availabel in Stage 2 studies for validation
1741 of Stage 1 association.

1742