View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Apollo

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

Anne E Justice[¥,1,2], Tugce Karaderi[¥,3,4], Heather M Highland[¥,1,5], Kristin L Young[¥,1], Mariaelisa Graff[¥,1], Yingchang Lu[¥,6,7,8], Valérie Turcot[9], Paul L Auer[10], Rebecca S Fine[11,12,13], Xiuqing Guo[14], Claudia Schurmann[7,8], Adelheid Lempradl[15], Eirini Marouli[16], Anubha Mahajan[3], Thomas W Winkler[17], Adam E Locke[18,19], Carolina Medina-Gomez[20,21], Tõnu Esko[11,13,22], Sailaja Vedantam[11,12,13], Ayush Giri[23], Ken Sin Lo[9], Tamuno Alfred[7], Poorva Mudgal[24], Maggie CY Ng[24,25], Nancy L Heard-Costa[26,27], Mary F Feitosa[28], Alisa K Manning[11,29,30], Sara M Willems[31], Suthesh Sivapalaratnam[30,32,33], Goncalo Abecasis[18], Dewan S Alam[34], Matthew Allison[35], Philippe Amouyel[36,37,38], Zorayr Arzumanyan[14], Beverley Balkau[39], Lisa Bastarache[40], Sven Bergmann[41,42], Lawrence F Bielak[43], Matthias Blüher[44,45], Michael Boehnke[18], Heiner Boeing[46], Eric Boerwinkle[47,48], Carsten A Böger[49], Jette Bork-Jensen[50], Erwin P Bottinger[7], Donald W Bowden[24,25,51], Ivan Brandslund[52,53], Linda Broer[21], Amber A Burt[54], Adam S Butterworth[55,56], Mark J Caulfield[16,57], Giancarlo Cesana[58], John C Chambers[59,60,61,62,63], Daniel I Chasman[11,64,65,66], Yii-Der Ida Chen[14], Rajiv Chowdhury[55], Cramer Christensen[67], Audrey Y Chu[27,65], Francis S Collins[68], James P Cook[69], Amanda J Cox[24,25,70], David S Crosslin[71], John Danesh[55,56,72,73], Paul IW de Bakker[74,75], Simon de Denus[9,76], Renée de Mutsert[77], George Dedoussis[78], Ellen W Demerath[79], Joe G Dennis[80], Josh C Denny[40], Emanuele Di Angelantonio[55,56], Marcus Dörr[81,82], Fotios Drenos[83,84], Marie-Pierre Dubé[9,85], Alison M Dunning[86], Douglas F Easton[80,86], Paul Elliott[87], Evangelos Evangelou[61,88], Aliki-Eleni Farmaki[78], Shuang Feng[18], Ele Ferrannini[89,90], Jean Ferrieres[91], Jose C Florez[11,29,30], Myriam Fornage[92], Caroline S Fox[27], Paul W Franks[93,94,95], Nele Friedrich[96], Wei Gan[3], Ilaria Gandin[97], Paolo Gasparini[98,99], Vilmantas Giedraitis[100], Giorgia Girotto[98,99], Mathias Gorski[17,49], Harald Grallert[101,102,103], Niels Grarup[50], Megan L Grove[47], Stefan Gustafsson[104], Jeff Haessler[105], Torben Hansen[50], Andrew T Hattersley[106], Caroline Hayward[107], Iris M Heid[17,108], Oddgeir L Holmen[109], G Kees Hovingh[110], Joanna MM Howson[55], Yao Hu[111], Yi-Jen Hung[112,113], Kristian Hveem[109,114], M Arfan Ikram[20,115,116], Erik

24 Ingelsson[104,117], Anne U Jackson[18], Gail P Jarvik[54,118], Yucheng Jia [14], Torben Jørgensen[119,120,121], Pekka
25 Jousilahti[122], Johanne M Justesen[50], Bratati Kahali[123,124,125,126], Maria Karaleftheri[127], Sharon LR Kardia[43],
26 Fredrik Karpe[128,129], Frank Kee[130], Hidetoshi Kitajima[3], Pirjo Komulainen[131,132,133], Jaspal S Kooner[60,62,63,134],
27 Peter Kovacs[44], Bernhard K Krämer[135], Kari Kuulasmaa[122], Johanna Kuusisto[136], Markku Laakso[136], Timo A
28 Lakka[131,132,133], David Lamparter[41,42], Leslie A Lange[137], Claudia Langenberg[31], Eric B Larson[54,138,139], Nanette
29 R Lee[140,141], Wen-Jane Lee[142,143], Terho Lehtimäki[144,145], Cora E Lewis[146], Huaixing Li[111], Jin Li[147], Ruifang Li-
30 Gao[77], Li-An Lin[92], Xu Lin[111], Lars Lind[148], Jaana Lindström[122], Allan Linneberg[121,149,150], Ching-Ti Liu[151],
31 Dajiang J Liu[152], Jian'an Luan[31], Leo-Pekka Lyytikäinen[144,145], Stuart MacGregor[153], Reedik Mägi[22], Satu
32 Männistö[122], Gaëlle Marenne[72], Jonathan Marten[107], Nicholas GD Masca[154,155], Mark I McCarthy[3,128,129],
33 Karina Meidtner[101,156], Evelin Mihailov[22], Leena Moilanen[157], Marie Moitry[158,159], Dennis O Mook-
34 Kanamori[77,160], Anna Morgan[98], Andrew P Morris[3,69], Martina Müller-Nurasyid[108,161,162], Patricia B
35 Munroe[16,57], Narisu Narisu[68], Christopher P Nelson[154,155], Matt Neville[128,129], Ioanna Ntalla[16], Jeffrey R
36 O'Connel[163], Katharine R Owen[128,129], Oluf Pedersen[50], Gina M Peloso[151], Craig E Pennell[164], Markus
37 Perola[122,165,166], James A Perry[163], John RB Perry[31], Tune H Pers[50,167], Ailith Pirie[86], Ozren Polasek[168,169], Olli
38 T Raitakari[170,171], Asif Rasheed[172], Chelsea K Raulerson[137], Rainer Rauramaa[131,132,133], Dermot F Reilly[173], Alex
39 P Reiner[105,174], Paul M Ridker[65,66,175], Manuel A Rivas[11,176], Neil R Robertson[3,128], Antonietta Robino[177], Igor
40 Rudan[169], Katherine S Ruth[178], Danish Saleheen[172,179], Veikko Salomaa[122], Nilesh J Samani[154,155], Pamela J
41 Schreiner[180], Matthias B Schulze[101,156], Robert A Scott[31], Marcelo P Segura-Lepe[61], Xueling Sim[18,181], Andrew
42 J Slater[182,183], Kerrin S Small[184], Blair H Smith[185,186], Jennifer A Smith[43], Lorraine Southam[3,72], Timothy D
43 Spector[184], Elizabeth K Speliotes[123,124,125], Kari Stefansson[187,188], Valgerdur Steinthorsdottir[187], Kathleen E
44 Stirrups[16,33], Konstantin Strauch[108,189], Heather M Stringham[18], Michael Stumvoll[44,45], Liang Sun[190,191],
45 Praveen Surendran[55], Karin MA Swart[192], Jean-Claude Tardif[9,85], Kent D Taylor[14], Alexander Teumer[193],
46 Deborah J Thompson[80], Gudmar Thorleifsson[187], Unnur Thorsteinsdottir[187,188], Betina H Thuesen[121], Anke
47 Tönjes[194], Mina Torres[195], Emmanouil Tsafantakis[196], Jaakko Tuomilehto[122,197,198,199], André G

48 Uitterlinden[20,21], Matti Uusitupa[200], Cornelia M van Duijn[20], Mauno Vanhala[201,202], Rohit Varma[195], Sita H
49 Vermeulen[203], Henrik Vestergaard[50,204], Veronique Vitart[107], Thomas F Vogt[205], Dragana Vuckovic[99], Lynne
50 E Wagenknecht[206], Mark Walker[207], Lars Wallentin[208], Feijie Wang[111], Carol A Wang[164], Shuai Wang[151],
51 Nicholas J Wareham[31], Helen R Warren[16,57], Dawn M Waterworth[209], Jennifer Wessel[210], Harvey D White[211],
52 Cristen J Willer[123,124,212], James G Wilson[213], Andrew R Wood[178], Ying Wu[137], Hanieh Yaghootkar[178], Jie Yao[14],
53 Laura M Yerges-Armstrong[163,214], Robin Young[55,215], Eleftheria Zeggini[72], Xiaowei Zhan[216], Weihua
54 Zhang[60,61], Jing Hua Zhao[31], Wei Zhao[179], He Zheng[111], Wei Zhou[123,124], M Carola Zillikens[20,21], CHD Exome+
55 Consortium, EPIC-CVD Consortium, ExomeBP Consortium, Global Lipids Genetic Consortium, GoT2D
56 Genes Consortium, InterAct, ReproGen Consortium, T2D-Genes Consortium, The MAGIC Investigators,
57 Fernando Rivadeneira[20,21], Ingrid B Borecki[28], John A Pospisilik[15], Panos Deloukas[16,217], Timothy M
58 Frayling[178], Guillaume Lettre[9,85], Karen L Mohlke[137], Jerome I Rotter[14], Zoltán Kutalik[42,218], Joel N
59 Hirschhorn[11,13,219], L Adrienne Cupples[Ł,27,151], Ruth JF Loos[Ł,7,8,220], *Kari E North[*,Ł,221], *Cecilia M Lindgren
60 [*,Ł,3,11,222]

61 ¥ These authors contributed equally to this work.

62 Ł These authors jointly supervised this work.

63 ***CORRESPONDING AUTHORS**

64 Prof. Kari North
65 Department of Epidemiology
66 University of North Carolina at Chapel Hill
67 137 East Franklin Street
68 Suite 306
69 Chapel Hill, NC 27514
70
71 Prof. Cecilia M Lindgren
72 The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery
73 University of Oxford
74 Roosevelt Drive

75 Oxford
76 OX3 7BN
77 United Kingdom
78 celi@well.ox.ac.uk

79 **AFFILIATIONS**

80 1. Department of Epidemiology, University of North Carolina, Chapel Hill, NC, 27514, USA
81 2. Biomedical and Translational Informatics, Geisinger Health, Danville, PA 17822
82 3. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
83 4. Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean
84 University, Famagusta, Cyprus
85 5. Human Genetics Center, The University of Texas School of Public Health, The University of Texas
86 MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, The University of
87 Texas Health Science Center at Houston, Houston, TX, 77030, USA
88 6. Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt
89 Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, 37203, USA
90 7. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount
91 Sinai, New York, NY, 10029, USA
92 8. The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount
93 Sinai, New York, NY, 10069, USA
94 9. Montreal Heart Institute, Universite de Montreal, Montreal, Quebec, H1T 1C8, Canada
95 10. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, 53201, USA
96 11. Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
97 12. Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA
98 13. Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston
99 Children's Hospital, Boston, MA, 02115, USA

100 14. Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical
101 Center, Torrance, CA, 90502, USA
102 15. Max Planck Institute of Immunobiology and Epigenetics, Freiburg, 79108, Germany
103 16. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen
104 Mary University of London, London, EC1M 6BQ, UK
105 17. Department of Genetic Epidemiology, University of Regensburg, Regensburg, D-93051, Germany
106 18. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor,
107 MI, 48109, USA
108 19. McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, 63108,
109 USA
110 20. Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
111 21. Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
112 22. Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia
113 23. Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health,
114 Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, 37203, USA
115 24. Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA
116 25. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine,
117 Winston-Salem, NC, 27157, USA
118 26. Department of Neurology, Boston University School of Medicine, Boston, MA, 02118, USA
119 27. NHLBI Framingham Heart Study, Framingham, MA, 01702, USA
120 28. Division of Statistical Genomics, Department of Genetics, Washington University School of
121 Medicine, St. Louis, MO, 63108, USA
122 29. Department of Medicine, Harvard University Medical School, Boston, MA, 02115, USA
123 30. Massachusetts General Hospital, Boston, MA, 02114, USA

124 31. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of
125 Metabolic Science, Cambridge, CB2 0QQ, UK
126 32. Department of Vascular Medicine, AMC, Amsterdam, 1105 AZ, The Netherlands
127 33. Department of Haematology, University of Cambridge, Cambridge, CB2 0PT, UK
128 34. School of Kinesiology and Health Science, Faculty of Health, York University, Toronto
129 35. Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA,
130 92093, USA
131 36. INSERM U1167, Lille, F-59019, France
132 37. Institut Pasteur de Lille, U1167, Lille, F-59019, France
133 38. Universite de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related
134 diseases, Lille, F-59019, France
135 39. INSERM U1018, Centre de recherche en Épidemiologie et Sante des Populations (CESP), Villejuif,
136 France
137 40. Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, 37203, USA
138 41. Department of Computational Biology, University of Lausanne, Lausanne, 1011, Switzerland
139 42. Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
140 43. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI,
141 48109, USA
142 44. IFB Adiposity Diseases, University of Leipzig, Leipzig, 04103, Germany
143 45. University of Leipzig, Department of Medicine, Leipzig, 04103, Germany
144 46. Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE),
145 Nuthetal, 14558, Germany
146 47. School of Public Health, Human Genetics Center, The University of Texas Health Science Center at
147 Houston, Houston, TX, 77030, USA

148 48. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030 USA

149 49. Department of Nephrology, University Hospital Regensburg, Regensburg, 93042, Germany

150 50. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical

151 Sciences, University of Copenhagen, Copenhagen, 2100, Denmark

152 51. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA

153 52. Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, 7100, Denmark

154 53. Institute of Regional Health Research, University of Southern Denmark, Odense, 5000, Denmark

155 54. Department of Medicine, University of Washington, Seattle, WA, 98195, USA

156 55. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care,

157 University of Cambridge, Cambridge, CB1 8RN, UK

158 56. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public

159 Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK

160 57. NIHR Barts Cardiovascular Research Unit, Barts and The London School of Medicine & Dentistry,

161 Queen Mary University of London, London, EC1M 6BQ, UK

162 58. Research Centre on Public Health, University of Milano-Bicocca, Monza, 20900, Italy

163 59. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232,

164 Singapore

165 60. Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, Middlesex,

166 UB1 3HW, UK

167 61. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,

168 London, W2 1PG, UK

169 62. Imperial College Healthcare NHS Trust, London, W12 0HS, UK

170 63. MRC-PHE Centre for Environment and Health, Imperial College London, London, W2 1PG, UK

171 64. Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA,
172 02115, USA
173 65. Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston, MA,
174 02215, USA
175 66. Harvard Medical School, Boston, MA, 02115, USA
176 67. Medical department, Lillebaelt Hospital, Vejle, 7100, Denmark
177 68. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute,
178 National Institutes of Health, Bethesda, MD, 20892, USA
179 69. Department of Biostatistics, University of Liverpool, Liverpool, L69 3GL, UK
180 70. Menzies Health Institute Queensland, Griffith University, Southport, QLD, Australia
181 71. Department of Biomedical Infomatics and Medical Education, University of Washington, Seattle,
182 WA, 98195, USA
183 72. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK
184 73. British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of
185 Cambridge, Cambridge, CB2 0QQ, UK
186 74. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht,
187 The Netherlands
188 75. Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht,
189 Utrecht, 3584 CX, The Netherlands
190 76. Faculty of Pharmacy, Universite de Montreal, Montreal, Quebec, H3T 1J4, Canada
191 77. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2300RC, The
192 Netherlands
193 78. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio
194 University, Athens, 17671, Greece

195 79. Division of Epidemiology & Community Health, School of Public Health, University of Minnesota,
196 Minneapolis, MN, 55454, USA
197 80. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care,
198 University of Cambridge, Cambridge, CB1 8RN, UK
199 81. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, 17475, Germany
200 82. DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, 17475,
201 Germany
202 83. Institute of Cardiovascular Science, University College London, London, WC1E 6JF, UK
203 84. MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of
204 Bristol, Bristol, BS8 2BN, UK
205 85. Department of Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, H3T
206 1J4, Canada
207 86. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge,
208 Cambridge, CB1 8RN, UK
209 87. Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health,
210 School of Public Health, Imperial College London, London, W2 1PG, UK
211 88. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina,
212 45110, Greece
213 89. CNR Institute of Clinical Physiology, Pisa, Italy
214 90. Department of Clinical & Experimental Medicine, University of Pisa, Italy
215 91. Toulouse University School of Medicine, Toulouse, TSA 50032 31059, France
216 92. Institute of Molecular Medicine, The University of Texas Health Science Center at Houston,
217 Houston, TX, 77030, USA

218 93. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University,
219 Malmo, SE-20502, Sweden
220 94. Department of Nutrition, Harvard School of Public Health, Boston, MA, 02115, USA
221 95. Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, 901
222 87, Sweden
223 96. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald,
224 Greifswald, 17475, Germany
225 97. Ilaria Gandin, Research Unit, AREA Science Park, Trieste, 34149, Italy
226 98. Department of Medical Sciences, University of Trieste, Trieste, 34137, Italy
227 99. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy
228 100. Geriatrics, Department of Public Health, Uppsala University, Uppsala, 751 85, Sweden
229 101. German Center for Diabetes Research, München-Neuherberg, 85764, Germany
230
231 102. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for
232 Environmental Health, Neuherberg, 85764, Germany
233 103. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research
234 Center for Environmental Health, Neuherberg, 85764, Germany
235 104. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory,
236 Uppsala University, Uppsala, 751 41, Sweden
237 105. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, 98109,
238 USA
239 106. University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
240 107. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh,
241 Edinburgh, EH4 2XU, UK

242 108. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for
243 Environmental Health, Neuherberg, 85764, Germany
244 109. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian
245 University of Science and Technology, Trondheim, 7600, Norway
246 110. AMC, Department of Vascular Medicine, Amsterdam, 1105 AZ, The Netherlands
247 111. Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes
248 for Biological Sciences, Chinese Academy of Sciences, University of the Chinese Academy of
249 Sciences, Shanghai, People's Republic of China, Shanghai, 200031, China
250 112. Division of Endocrinology and Metabolism, Department of Internal Medicine, Tri-Service General
251 Hospital, Taipei, Taiwan 114, Taiwan
252 113. School of Medicine, National Defense Medical Center, Taipei, Taiwan 114, Taiwan
253 114. HUNT Research center, Department of Public Health, Norwegian University of Science and
254 Technology, Levanger, 7600, Norway
255 115. Department of Neurology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
256 116. Department of Radiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
257 117. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of
258 Medicine, Stanford, CA, 943 05, USA
259 118. Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA
260 119. Faculty of medicine, Aalborg University, Aalborg, DK-9000, Denmark
261 120. Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen,
262 Copenhagen, 2200, Denmark
263 121. Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, DK-2600,
264 Denmark
265 122. National Institute for Health and Welfare, Helsinki, FI-00271, Finland

266 123. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor,
267 MI, 48109, USA
268 124. Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109, USA
269 125. Division of Gastroenterology, University of Michigan, Ann Arbor, MI, 48109, USA
270 126. Centre for Brain Research, Indian Institute of Science, Bangalore 560012, India
271 127. Echinos Medical Centre, Echinos, Greece
272 128. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine,
273 University of Oxford, Oxford, OX3 7LE, UK
274 129. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK
275 130. UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast, UK,
276 BT12 6BJ, UK
277 131. Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise
278 Medicine, Kuopio, 70100, Finland
279 132. Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus,
280 70210, Finland
281 133. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio,
282 Finland
283 134. National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus,
284 London, W12 0NN, UK
285 135. University Medical Centre Mannheim, 5th Medical Department, University of Heidelberg,
286 Mannheim, 68167, Germany
287 136. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio
288 University Hospital, Kuopio, 70210, Finland
289 137. Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA

290 138. Kaiser Permanente Washington Health Research Institute Seattle WA 98101
291 139. Department of Health Services, University of Washington, Seattle WA 98101
292 140. Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, 6000,
293 Philippines
294 141. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, 6000,
295 Philippines
296 142. Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan 407,
297 Taiwan
298 143. Department of Social Work, Tunghai University, Taichung, Taiwan
299 144. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, 33521, Finland
300 145. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of
301 Medicine and Life Sciences, University of Tampere, Tampere 33014, Finland
302 146. Division of Preventive Medicine University of Alabama at Birmingham, Birmingham, AL 35205,
303 USA
304 147. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of
305 Medicine, Palo Alto, CA, 94304, USA
306 148. Uppsala University, Uppsala, 75185, Sweden
307 149. Department of Clinical Experimental Research, Rigshospitalet, Copenhagen, DK-2200, Denmark
308 150. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of
309 Copenhagen, Copenhagen, 2200, Denmark
310 151. Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA
311 152. Department of Public Health Sciences, Institute for Personalized Medicine, the Pennsylvania State
312 University College of Medicine, Hershey, PA, 17033, USA
313 153. QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia

314 154. Department of Cardiovascular Sciences, Univeristy of Leicester, Glenfield Hospital, Leicester, LE3
315 9QP, UK
316 155. NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP,
317 UK
318 156. Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-
319 Rehbruecke (DIfE), Nuthetal, 14558, Germany
320 157. Department of Medicine, Kuopio University Hospital, Kuopio, 70210, Finland
321 158. Department of Epidemiology and Public Health, University of Strasbourg, Strasbourg, F-67085,
322 France
323 159. Department of Public Health, University Hospital of Strasbourg, Strasbourg, F-67081, France
324 160. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, 2300RC,
325 The Netherlands
326 161. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universitat,
327 Munich, 81377, Germany
328 162. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich,
329 80802, Germany
330 163. Program for Personalized and Genomic Medicine, Department of Medicine, University of
331 Maryland School of Medicine, Baltimore, MD, 21201, US
332 164. School of Women's and Infants' Health, The University of Western Australia, Perth, Western
333 Australia, 6009, Australia
334 165. University of Helsinki, Institute for Molecular Medicine (FIMM) and Diabetes and Obesity
335 Research Program, Helsinki, FI00014, Finland
336 166. University of Tartu, Estonian Genome Center, Tartu, Estonia, Tartu, 51010, Estonia
337 167. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, 2200, Denmark

338 168. School of Medicine, University of Split, Split, 21000, Croatia

339 169. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics,

340 University of Edinburgh, Edinburgh, EH8 9AG, UK

341 170. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, 20521,

342 Finland

343 171. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku,

344 20520, Finland

345 172. Centre for Non-Communicable Diseases, Karachi, Pakistan

346 173. Merck, Sharp & Dohme, Genetics and Pharmacogenomics, Boston, MA, 02115, USA

347 174. Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA

348 175. Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School,

349 Boston, MA, 02115, USA

350 176. Nuffield Department of Clinical Medicine, Oxford, OX3 7BN, UK

351 177. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, 34137, Italy

352 178. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2

353 5DW, UK

354 179. Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of

355 Pennsylvania, Philadelphia, PA, 19104, USA

356 180. Division of Epidemiology & Community Health University of Minnesota, Minneapolis, MN, 55454,

357 USA

358 181. Saw Swee Hock School of Public Health, National University Health System, National University of

359 Singapore, Singapore 117549, Singapore

360 182. Genetics, Target Sciences, GlaxoSmithKline, Research Triangle Park, NC, 27709, US

361 183. OmicSoft a QIAGEN Company, Cary, NC, 27513, US

362 184. Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH,
363 UK
364 185. Division of Population Health Sciences, Ninewells Hospital and Medical School, University of
365 Dundee, Dundee, UK
366 186. Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh,
367 Edinburgh, EH4 2XU, UK
368 187. deCODE Genetics/Amgen inc., Reykjavik, 101, Iceland
369 188. Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland
370 189. Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, 81377, Germany
371 190. Biodemography of Aging Research Unit, Social Science Research Institute, Duke University,
372 Durham, NC, 27708, USA
373 191. Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland
374 192. VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam, 1007
375 MB, The Netherlands
376 193. Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany
377 194. Center for Pediatric Research, Department for Women's and Child Health, University of Leipzig,
378 Leipzig, 04103, Germany
379 195. USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University
380 of Southern California, Los Angeles, CA, 90033, USA
381 196. Anogia Medical Centre, Anogia, Greece
382 197. Centre for Vascular Prevention, Danube-University Krems, Krems, 3500, Austria
383 198. Dasman Diabetes Institute, Dasman, 15462, Kuwait
384 199. Diabetes Research Group, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

385 200. Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, 70210,
386 Finland
387 201. Central Finland Central Hospital, Jyvaskyla, 40620, Finland
388 202. University of Eastern Finland, Kuopio, 70210, Finland
389 203. Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, 6500 HB,
390 The Netherlands
391 204. Steno Diabetes Center Copenhagen, Gentofte, 2800, Denmark
392 205. Merck, Sharp & Dohme, Cardiometabolic Disease, Kenilworth, NJ, 07033, USA
393 206. Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, 27157,
394 USA
395 207. Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, NE2 4HH, UK
396 208. Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala
397 University, Uppsala, 752 37, Sweden
398 209. Genetics, Target Sciences, GlaxoSmithKline, King of Prussia, PA, US
399 210. Departments of Epidemiology & Medicine, Diabetes Translational Research Center, Fairbanks
400 School of Public Health & School of Medicine, Indiana University, Indiana, IN, 46202, USA
401 211. Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland,
402 New Zealand
403 212. Department of Human Genetics, University of Michigan, Ann Arbor, MI, 48109, USA
404 213. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS,
405 39216, USA
406 214. GlaxoSmithKline, King of Prussia, PA, 19406, USA
407 215. University of Glasgow, Glasgow, G12 8QQ, UK

408 216. Department of Clinical Sciences, Quantitative Biomedical Research Center, Center for the
409 Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, 75390,
410 USA
411 217. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-
412 HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia
413 218. Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, 1010,
414 Switzerland
415 219. Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, 02115, USA
416 220. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai,
417 New York, NY, 10069, USA
418 221. Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, NC, 27514,
419 USA
420 222. Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of
421 Oxford, Oxford, OX3 7BN, UK
422

423 **ABSTRACT**

424 Body fat distribution is a heritable risk factor for a range of adverse health consequences, 425 including hyperlipidemia and type 2 diabetes. To identify protein-coding variants associated with body fat 426 distribution, assessed by waist-to-hip ratio adjusted for body mass index, we analyzed 228,985 predicted 427 coding and splice site variants available on exome arrays in up to 344,369 individuals from five major 428 ancestries for discovery and 132,177 independent European-ancestry individuals for validation. We 429 identified 15 common (minor allele frequency, MAF ≥ 5%) and 9 low frequency or rare (MAF < 5%) coding 430 variants that have not been reported previously. Pathway/gene set enrichment analyses of all associated 431 variants highlight lipid particle, adiponectin level, abnormal white adipose tissue physiology, and bone 432 development and morphology as processes affecting fat distribution and body shape. Furthermore, the 433 cross-trait associations and the analyses of variant and gene function highlight a strong connection to 434 lipids, cardiovascular traits, and type 2 diabetes. In functional follow-up analyses, specifically in *Drosophila* 435 RNAi-knockdown crosses, we observed a significant increase in the total body triglyceride levels for two 436 genes (*DNAH10* and *PLXND1*). By examining variants often poorly tagged or entirely missed by genome- 437 wide association studies, we implicate novel genes in fat distribution, stressing the importance of 438 interrogating low-frequency and protein-coding variants.

439
440
441
442
443
444

445 Body fat distribution, as assessed by waist-to-hip ratio (WHR), is a heritable trait and a well-
446 established risk factor for adverse metabolic outcomes[1-6]. A high WHR often indicates a large presence
447 of intra-abdominal fat whereas a low WHR is correlated with a greater accumulation of gluteofemoral
448 fat. Lower values of WHR have been consistently associated with lower risk of cardiometabolic diseases
449 like type 2 diabetes (T2D)[7,8], or differences in bone structure and gluteal muscle mass[9]. These
450 epidemiological associations are consistent with the results of our previously reported genome-wide
451 association study (GWAS) of 49 loci associated with WHR (after adjusting for body mass index,
452 WHRadjBMI)[10]. Notably, a genetic predisposition to higher WHRadjBMI is associated with increased risk
453 of T2D and coronary heart disease (CHD), and this association appears to be causal[9].

454 More recently, large-scale genetic studies have identified ~125 common loci for central obesity,
455 primarily non-coding variants of relatively modest effect, for different measures of body fat distribution[10-
456 16]. Large scale interrogation of both common (minor allele frequency [MAF]≥5%) and low frequency or
457 rare (MAF<5%) coding and splice site variation may lead to additional insights into the genetic and
458 biological etiology of central obesity by narrowing in on causal genes contributing to trait variance. Thus,
459 we set out to identify protein-coding and splice site variants associated with WHRadjBMI using exome
460 array data and to explore their contribution to variation in WHRadjBMI through multiple follow-up
461 analyses.

462 **RESULTS**

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

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

468 Asian (N=6,800) and Hispanic/Latino (N=4,075) descent, genotyped with an ExomeChip array
469 (**Supplementary Tables 1-3).** For stage 2, we assessed 70 suggestively significant ($P<2x10^{-6}$) variants from
470 stage 1 in two independent cohorts from the United Kingdom [UK Biobank (UKBB), N=119,572] and
471 Iceland (deCODE, N=12,605) (**Online Methods, Supplementary Data 1-3**) for a total stage 1+2 sample size
472 of 476,546 (88% European). Variants were considered statistically significant in the total meta-analyzed
473 sample (stage 1+2) when they achieved a significance threshold of $P<2x10^{-7}$ after Bonferroni correction
474 for multiple testing (0.05/246,328 variants tested). Of the 70 variants brought forward, two common and
475 five rare variants were not available in either Stage 2 study (**Tables 1-2, Supplementary Data 1-3**). Thus,
476 we require $P<2x10^{-7}$ in Stage 1 for significance. Variants are considered novel if they were greater than
477 one megabase (Mb) from a previously-identified WHRadjBMI lead SNP[10-16].

478 In stages 1 and 2 combined all ancestry meta-analyses, we identified 48 coding variants (16 novel)
479 across 43 genes, 47 identified assuming an additive model, and one more variant under a recessive model
480 (**Table 1, Supplementary Figures 1-4**). Due to the possible heterogeneity introduced by combining
481 multiple ancestries[17], we also performed a European-only meta-analysis. Here, four additional coding
482 variants were significant (three novel) assuming an additive model (**Table 1**, **Supplementary Figures 5-8**).
483 Of these 52 significant variants (48 from the all ancestry and 4 from the European-only analyses), eleven
484 were of low frequency, including seven novel variants in *RAPGEF3*, *FGFR2, R3HDML, HIST1H1T, PCNXL3,*
485 *ACVR1C,* and *DARS2*. These low frequency variants tended to display larger effect estimates than any of
486 the previously reported common variants (**Figure 2**)[10]. In general, variants with MAF<1% had effect sizes
487 approximately three times greater than those of common variants (MAF>5%). Although, we cannot rule
488 out the possibility that additional rare variants with smaller effects sizes exist that, despite our ample
489 sample size, we are still underpowered to detect (See estimated 80% power in **Figure 2**). However, in the
490 absence of common variants with similarly large effects, our results point to the importance of
491 investigating rare and low frequency variants to identify variants with large effects (**Figure 2**).

492 Given the established differences in the genetic underpinnings between sexes for
493 WHRadjBMI[10,11], we also performed sex-stratified analyses and report variants that were array-wide
494 significant ($P<2x10^{-7}$) in at least one sex stratum and exhibit significant sex-specific effects ($P_{sexhet}<7.14x10^{-4}$
495 , see **Online Methods**). We found four additional novel variants that were not identified in the sex-
496 combined meta-analyses (in *UGGT2* and *MMP14* for men only; and *DSTYK* and *ANGPTL4* for women only)
497 (**Table 2, Supplementary Figures 9-15**). Variants in *UGGT2* and *ANGPTL4* were of low frequency
498 (MAF_{men}=0.6% and MAF_{women}=1.9%, respectively). Additionally, 14 variants from the sex-combined meta-
499 analyses displayed stronger effects in women, including the novel, low frequency variant in *ACVR1C*
500 (rs55920843, MAF=1.1%, **Supplementary Figure 4**). Overall, 19 of the 56 variants (32%) identified across
501 all meta-analyses (48 from all ancestry, 4 from European-only and 4 from sex-stratified analyses) showed
502 significant sex-specific effects on WHRadjBMI (**Figure 1**): 16 variants with significantly stronger effects in
503 women, and three in men (**Figure 1**).

504 In summary, we identified 56 array-wide significant coding variants ($P<2.0x10^{-7}$); 43 common (14
505 novel) and 13 low frequency or rare variants (9 novel). For all 55 significant variants from the additive
506 model (47 from all ancestry, 4 from European-only, and 4 from sex-specific analyses), we examined
507 potential collider bias[18,19], i.e. potential bias in effect estimates caused by adjusting for a correlated and
508 heritable covariate like BMI, for the relevant sex stratum and ancestry. We corrected each of the variant
509 - WHRadjBMI associations for the correlation between WHR and BMI and the correlation between the
510 variant and BMI (**Online Methods, Supplementary Table 7, Supplementary Note 1**). Overall, 51 of the 55
511 additive model variants were robust against collider bias[18,19] across all primary and secondary meta-
512 analyses. Of the 55, 25 of the WHRadjBMI variants from the additive model were nominally associated
513 with BMI ($P_{BMI}<0.05$), yet effect sizes changed little after correction for potential biases (15% change in
514 effect estimate on average). For 4 of the 55 SNPs (rs141845046, rs1034405, rs3617, rs9469913, **Table 1**),
515 the association with WHRadjBMI appears to be attenuated following correction ($P_{corrected}> 9x10^{-4}$,

516 0.05/55), including one novel variant, rs1034405 in *C3orf18*. Thus, these 4 variants warrant further
517 functional investigations to quantify their impact on WHR, as a true association may still exist, although
518 the effect may be slightly overestimated in the current analysis.

519 Using stage 1 meta-analysis results, we then aggregated low frequency variants across genes and
520 tested their joint effect with both SKAT and burden tests[20] (**Supplementary Table 8**, **Online Methods**). We
521 identified five genes that reached array-wide significance (P<2.5x10^{-6}, 0.05/16,222 genes tested),
522 *RAPGEF3*, *ACVR1C*, *ANGPTL4*, *DNAI1*, and *NOP2*. However, while all genes analyzed included more than
523 one variant, none remained significant after conditioning on the single variant with the most significant
524 p-value. We identified variants within *RAPGEF3*, *ACVR1C*, *ANGPTL4* that reached suggestive significance
525 in Stage 1 and chip-wide significance in stage 1+2 for one or more meta-analyses (**Tables 1 and 2**);
526 however, we did not identify any significant variants for *DNAI1* and *NOP2.* While neither of these genes
527 had a single variant that reached chip-wide significance, they each had variants with nearly significant
528 results (*NOP2*: P=3.69x10^{-5}, *DNAI1*: 4.64x10^{-5}). Combined effects with these single variants and others in
529 LD within the gene likely drove the association in our aggregate gene-based tests, but resulted in non-
530 significance following conditioning on the top variant. While our results suggest these associations are
531 driven by a single variant, each gene may warrant consideration in future investigations.

532

533 **Conditional analyses**

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

540 $P_{conditional}$=4.37x10^{-23} in the combined sexes, and $P_{conditional}$=2.4x10^{-20} in women) and rs139745911 in

541 *KIAA0408* (MAF=0.9%, $P_{conditional}$=3.68x10^{-11} in the combined sexes, and $P_{conditional}$=1.46x10^{-11} in women;

542 **Figure 3A**).

543 Further, 33 of our significant variants are within one Mb of previously identified GWAS tag SNPs

544 for WHRadjBMI. We again used approximate joint conditional analysis to test for independence in the

545 stage 1 meta-analysis dataset and obtained further complementary evidence from the UKBB dataset

546 where necessary (**Online Methods**). We identified one coding variant representing a novel independent

547 signal in a known locus [*RREB1*; stage1 meta-analysis, rs1334576, EAF = 0.44, $P_{conditional}$= 3.06x10^{-7},

548 (**Supplementary Table 5, Figure 3 [B])**; UKBB analysis, rs1334576, *RREB1*, $P_{conditional}$= 1.24x10^{-8},

549 (**Supplementary Table 6**) in the sex-combined analysis.

550 In summary, we identified a total of 56 WHRadjBMI-associated coding variants in 41 independent

551 association signals. Of these 41 independent association signals, 24 are new or independent of known

552 GWAS-identified tag SNPs (either >1MB +/- or array-wide significant following conditional analyses)

553 (**Figure 1**). Thus, bringing our total to 15 common and 9 low-frequency or rare novel variants following

554 conditional analyses. The remaining non-GWAS-independent variants may assist in narrowing in on the

555 causal variant or gene underlying these established association signals.

556 Gene set and pathway enrichment analysis

557 To determine if the significant coding variants highlight novel biological pathways and/or provide

558 additional support for previously identified biological pathways, we applied two complementary pathway

559 analysis methods using the EC-DEPICT (ExomeChip Data-driven Expression Prioritized Integration for

560 Complex Traits) pathway analysis tool,[21,22] and PASCAL[23] (**Online Methods**). While for PASCAL all variants

561 were used, in the case of EC-DEPICT, we examined 361 variants with suggestive significance ($P<5x10^{-4}$)[10,17]

562 from the combined ancestries and combined sexes analysis (which after clumping and filtering became

563 101 lead variants in 101 genes). We separately analyzed variants that exhibited significant sex-specific

564 effects (P_{sexhet}<5x10^{-4}).

565 The sex-combined analyses identified 49 significantly enriched gene sets (FDR<0.05) that grouped

566 into 25 meta-gene sets (**Supplementary Note 2, Supplementary Data 4-5**). We noted a cluster of meta-

567 gene sets with direct relevance to metabolic aspects of obesity ("enhanced lipolysis," "abnormal glucose

568 homeostasis," "increased circulating insulin level," and "decreased susceptibility to diet-induced

569 obesity"); we observed two significant adiponectin-related gene sets within these meta-gene sets. While

570 these pathway groups had previously been identified in the GWAS DEPICT analysis (**Figure 4**), many of the

571 individual gene sets within these meta-gene sets were not significant in the previous GWAS analysis, such

572 as "insulin resistance," "abnormal white adipose tissue physiology," and "abnormal fat cell morphology"

573 **(Supplementary Data 4**, **Figure 4**, **Supplementary Figure 16a**), but represent similar biological

574 underpinnings implied by the shared meta-gene sets. Despite their overlap with the GWAS results, these

575 analyses highlight novel genes that fall outside known GWAS loci, based on their strong contribution to

576 the significantly enriched gene sets related to adipocyte and insulin biology (e.g. *MLXIPL, ACVR1C, and*

577 *ITIH5)* (**Figure 4**).

578 To focus on novel findings, we conducted pathway analyses after excluding variants from previous

579 WHRadjBMI analyses[10] (**Supplemental Note 2**). Seventy-five loci/genes were included in the EC-DEPICT

580 analysis, and we identified 26 significantly enriched gene sets (13 meta-gene sets). Here, all but one gene

581 set, "lipid particle size", were related to skeletal biology. This result likely reflects an effect on the pelvic

582 skeleton (hip circumference), shared signaling pathways between bone and fat (such as TGF-beta) and

583 shared developmental origin[24] (**Supplementary Data 5**, **Supplementary Figure 16b**). Many of these

584 pathways were previously found to be significant in the GWAS DEPICT analysis; these findings provide a

585 fully independent replication of their biological relevance for WHRadjBMI.

586 We used PASCAL (**Online Methods**) to further distinguish between enrichment based on *coding-*
587 *only* variant associations (this study) and *regulatory-only* variant associations (up to 20 kb upstream of the
588 gene from a previous GIANT study[10]). For completeness, we also compared the coding pathways to those
589 that could be identified in the total previous GWAS effort (using both *coding and regulatory* variants) by
590 PASCAL. The analysis revealed 116 significantly enriched coding pathways (FDR<0.05; **Supplementary**
591 **Table 9**). In contrast, a total of 158 gene sets were identified in the coding+regulatory analysis that
592 included data from the previous GIANT waist GWAS study. Forty-two gene sets were enriched in both
593 analyses. Thus, while we observed high concordance in the -log10 (p-values) between ExomeChip and
594 GWAS gene set enrichment (Pearson's r (coding *vs* regulatory only) = 0.38, $P<10^{-300}$; Pearson's r (coding *vs*
595 coding+regulatory) = 0.51, $P<10^{-300}$), there are gene sets that seem to be enriched *specifically* for variants
596 in coding regions (e.g., decreased susceptibility to diet-induced obesity, abnormal skeletal morphology)
597 or unique to variants in regulatory regions (e.g. transcriptional regulation of white adipocytes)
598 (**Supplementary Figure 17**).

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

605 **Cross-trait associations**

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

610 disease.[22,25-29] We found that variants in *STAB1* and *PLCB3* display the greatest number of significant cross-
611 trait associations, each associating with seven different traits ($P<9.8x10^{-4}$, 0.05/51 variants tested). Of
612 note, these two genes cluster together with *RSPO3*, *DNAH10*, *MNS1*, *COBLL1*, *CCDC92*, and *ITIH3*
613 (**Supplementary Data 6**, **Supplementary Figure 19**). The WHR-increasing alleles in this cluster of variants
614 exhibit a pattern of increased cardiometabolic risk (e.g. increased fasting insulin [FI], two-hour glucose
615 [TwoHGlu], and triglycerides [TG]; and decreased high-density lipoprotein cholesterol [HDL]), but also
616 decreased BMI. This phenomenon, where variants associated with lower BMI are also associated with
617 increased cardiometabolic risk, has been previously reported.[30-36]. A recent Mendelian Randomization
618 (MR) analysis of the relationship between central adiposity (measured as WHRadjBMI) and
619 cardiometabolic risk factors found central adiposity to be causal.[9] Using 48 WHR-increasing variants
620 reported in the recent GIANT analysis[10] to calculate a polygenic risk score, Emdin *et al*. found that a 1 SD
621 increase in genetic risk of central adiposity was associated with higher total cholesterol, triglyceride levels,
622 fasting insulin and two-hour glucose, and lower HDL – all indicators of cardiometabolic disease, and also
623 associated with a 1 unit decrease in BMI[9].

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

633 Given the established correlation between total body fat percentage and WHR (*R*= 0.052 to
634 0.483)[39-41], we examined the association of our top exome variants with both total body fat percentage
635 (BF%) and truncal fat percentage (TF%) available in a sub-sample of up to 118,160 participants of UKBB
636 (**Supplementary Tables 10-11**). Seven of the common novel variants were significantly associated
637 (*P*<0.001, 0.05/48 variants examined) with both BF% and TF% in the sexes-combined analysis (*COBLL1*,
638 *UHRF1BP1*, *WSCD2*, *CCDC92*, *IFI30*, *MPV17L2*, *IZUMO1*). Only one of our tag SNPs, rs7607980 in *COBLL1*,
639 is nearby a known ~~total body fat percentage~~BF% GWAS locus (rs6738627; R^2=0.1989, distance=6751 bp,
640 with our tag SNP)[42]. Two additional variants, rs62266958 in *EFCAB12* and rs224331 in *GDF5*, were
641 significantly associated with TF% in the women-only analysis. Of the nine SNPs associated with at least
642 one of these two traits, all variants displayed much greater magnitude of effect on TF% compared to BF%
643 (**Supplementary Figure 20**).

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

651 **Genetic architecture of WHRadjBMI coding variants**

652 We used summary statistics from our stage 1 results to estimate the phenotypic variance
653 explained by ExomeChip coding variants. We calculated the variance explained by subsets of SNPs across
654 various significance thresholds (P< $2x10^{-7}$ to 0.2) and conservatively estimated using only independent tag
655 SNPs (**Supplementary Table 12**, **Online Methods**, and **Supplementary Figure 22)**. The 22 independent
656 significant coding SNPs in stage 1 account for 0.28% of phenotypic variance in WHRadjBMI. For

657 independent variants that reached suggestive significance in stage 1 ($P<2x10^{-6}$), 33 SNPs explain 0.38% of
658 the variation; however, the 1,786 independent SNPs with a liberal threshold of P<0.02 explain 13 times
659 more variation (5.12%). While these large effect estimates may be subject to winner's curse, for array-
660 wide significant variants, we detected a consistent relationship between effect magnitude and MAF in our
661 stage 2 analyses in UK Biobank and deCODE (**Supplementary Data 1-3**). Notably, the Exomechip coding
662 variants explained less of the phenotypic variance than in our previous GIANT investigation, wherein 49
663 significant SNPs explained 1.4% of the variance in WHRadjBMI. When considering all coding variants on
664 the ExomeChip in men and women together, 46 SNPs with a $P<2x10^{-6}$ and 5,917 SNPs with a P<0.02 explain
665 0.51% and 13.75% of the variance in WHRadjBMI, respectively. As expected given the design of the
666 ExomeChip, the majority of the variance explained is attributable to rare and low frequency coding
667 variants (independent SNPs with MAF<1% and MAF<5% explain 5.18% and 5.58%, respectively). However,
668 for rare and low frequency variants, those that passed significance in stage 1 explain only 0.10% of the
669 variance in WHRadjBMI. As in **Figure 2**, these results also indicate that there are additional coding variants
670 associated with WHRadjBMI that remain to be discovered, particularly rare and low frequency variants
671 with larger effects than common variants. Due to observed differences in association strength between
672 women and men, we estimated variance explained for the same set of SNPs in women and men
673 separately. As observed in previous studies[10], there was significantly ($P_{RsqDiff}<0.002=0.05/21$, Bonferroni-
674 corrected threshold) more variance explained in women compared to men at each significance threshold
675 considered (differences ranged from 0.24% to 0.91%).

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

681 low frequency variants (MAF ≤ 1%) with larger effect sizes (effect size > 0.90) were included in the

682 penetrance analysis using World Health Organization (WHO- obese women WHR>0.85 and obese men

683 WHR>0.90) WHR cut-offs for central obesity. Of these, one SNV (rs55920843-*ACVR1C*; $P_{sex\text{-}combined}=9.25\times10^{-5}$; $P_{women}=4.85\times10^{-5}$) showed a statistically significant difference in the number of carriers and non-carriers

684

685 of the minor allele when the two strata were compared (sex-combined obese carriers=2.2%; non-obese

686 carriers=2.6%; women obese carriers=2.1%; non-obese women carriers=2.6% (**Supplementary Table 13,**

687 **Supplementary Figure 23**). These differences were significant in women, but not in men ($P_{men}<5.5\times10^{-3}$

688 after Bonferroni correction for 9 tests) and agree with our overall meta-analysis results, where the minor

689 allele (G) was significantly associated with lower WHRadjBMI in women only (**Tables 1 and 2**).

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

691 ***Drosophila Knockdown***

692 Considering the genetic evidence of adipose and insulin biology in determining body fat

693 distribution[10], and the lipid signature of the variants described here, we examined whole-body

694 triglycerides levels in adult *Drosophila*, a model organism in which the fat body is an organ functionally

695 analogous to mammalian liver and adipose tissue and triglycerides are the major source of fat storage [45].

696 Of the 51 genes harboring our 56 significantly associated variants, we identified 27 with *Drosophila*

697 orthologues for functional follow-up analyses. In order to prioritize genes for follow-up, we selected genes

698 with large changes in triglyceride storage levels (> 20% increase or > 40% decrease, as chance alone is

699 unlikely to cause changes of this magnitude, although some decrease is expected) after considering each

700 corresponding orthologue in an existing large-scale screen for adipose with ≤2 replicates per knockdown

701 strain.[45] Two orthologues, for *PLXND1* and *DNAH10*, from two separate loci met these criteria. For these

702 two genes, we conducted additional knockdown experiments with ≥5 replicates using tissue-specific

703 drivers (fat body [cg-Gal4] and neuronal [elav-Gal4] specific RNAi-knockdowns) (**Supplementary Table**

704 **14**). A significant (P<0.025, 0.05/2 orthologues) increase in the total body triglyceride levels was observed

705 in *DNAH10* orthologue knockdown strains for both the fat body and neuronal drivers. However, only the
706 neuronal driver knockdown for *PLXND1* produced a significant change in triglyceride storage. *DNAH10*
707 and *PLXND1* both lie within previous GWAS identified regions. Adjacent genes have been highlighted as
708 likely candidates for the *DNAH10* association region, including *CCDC92* and *ZNF664* based on eQTL
709 evidence. However, our fly knockdown results support *DNAH10* as the causal genes underlying this
710 association. Of note, rs11057353 in *DNAH10* showed suggestive significance after conditioning on the
711 known GWAS variants in nearby *CCDC92* (sex-combined $P_{conditional}$=7.56x10^{-7}; women-only rs11057353
712 $P_{conditional}$= 5.86x10^{-7}, **Supplementary Table 6**; thus providing some evidence of multiple causal
713 variants/genes underlying this association signal. Further analyses are needed to determine whether the
714 implicated coding variants from the current analysis are the putatively functional variants, specifically how
715 these variants affect transcription in and around these loci, and exactly how those effects alter biology of
716 relevant human metabolic tissues.

717 ***eQTL Lookups***

718 To gain a better understanding of the potential functionality of novel and low frequency variants,
719 we examined the *cis*-association of the identified variants with expression level of nearby genes in
720 subcutaneous adipose tissue, visceral omental adipose tissue, skeletal muscle and pancreas from GTEx[46],
721 and assessed whether the exome and eQTL associations implicated the same signal (**Online Methods**,
722 **Supplementary Data 9, Supplementary Table 15**). The lead exome variant was associated with expression
723 level of the coding gene itself for *DAGLB, MLXIPL, CCDC92, MAPKBP1, LRRC36* and *UQCC1*. However, at
724 three of these loci (*MLXIPL, MAPKBP1,* and *LRRC36)*, the lead exome variant is also associated with
725 expression level of additional nearby genes, and at three additional loci, the lead exome variant is only
726 associated with expression level of nearby genes (*HEMK1* at *C3orf18; NT5DC2, SMIM4* and *TMEM110* at
727 *STAB1/ITIH3;* and *C6orf106* at *UHRF1BP1*). Although detected with a missense variant, these loci are also

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

730 Some of the coding genes implicated by eQTL analyses are known to be involved in adipocyte
731 differentiation or insulin sensitivity: e. g. for *MLXIPL*, the encoded carbohydrate responsive element
732 binding protein is a transcription factor, regulating glucose-mediated induction of *de novo* lipogenesis in
733 adipose tissue, and expression of its *beta*-isoform in adipose tissue is positively correlated with adipose
734 insulin sensitivity[47,48]. For *CCDC92,* the reduced adipocyte lipid accumulation upon knockdown confirmed
735 the involvement of its encoded protein in adipose differentiation[49].

736 ***Biological Curation***

737 To gain further insight into the possible functional role of the identified variants, we conducted
738 thorough searches of the literature and publicly available bioinformatics databases (**Supplementary Data**
739 **10-11**, **Box 1**, **Online Methods**). Many of our novel low frequency variants are in genes that are intolerant
740 of nonsynonymous mutations (e.g. *ACVR1C*, *DARS2*, *FGFR2*; ExAC Constraint Scores >0.5). Like previously
741 identified GWAS variants, several of our novel coding variants lie within genes that are involved in glucose
742 homeostasis (e.g. *ACVR1C*, *UGGT2*, *ANGPTL4*), angiogenesis (*RASIP1*), adipogenesis (*RAPGEF3*), and lipid
743 biology (*ANGPTL4*, *DAGLB*) (**Supplementary Data 10**, **Box 1**).

744

745 **DISCUSSION**

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

751 underlying etiology. While, due to their rarity, these coding variants only explain a small proportion of the
752 trait variance at a population level, they may, given their predicted role, be more functionally tractable
753 than non-coding variants and have a critical impact at the individual and clinical level. For instance, the
754 association between a low frequency variant (rs11209026; R381Q; MAF<5% in ExAC) located in the *IL23R*
755 gene and multiple inflammatory diseases (such as psoriasis[50], rheumatoid arthritis[51], ankylosing
756 spondylitis[52], and inflammatory bowel diseases[53]) led to the development of new therapies, targeting *IL23*
757 and *IL12* in the same pathway (reviewed in [54-56]). Thus, we are encouraged that our associated low
758 frequency coding variants displayed large effect sizes; all but one of the nine novel low frequency variants
759 had an effect size larger than the 49 SNPs reported in Shungin *et al.* 2015, and some of these effect sizes
760 were up to 7-fold larger than those previously reported for GWAS. This finding mirrors results for other
761 cardiometabolic traits[57], and suggests variants of possible clinical significance with even larger effect and
762 lower frequency variants will likely be detected through larger additional genome-wide scans of many
763 more individuals.

764 We continue to observe sexual dimorphism in the genetic architecture of WHRadjBMI[11]. Overall,
765 we identified 19 coding variants that display significant sex differences, of which 16 (84%) display larger
766 effects in women compared to men. Of the variants outside of GWAS loci, we reported three (two with
767 MAF<5%) that show a significantly stronger effect in women and two (one with MAF<5%) that show a
768 stronger effect in men. Additionally, genetic variants continue to explain a higher proportion of the
769 phenotypic variation in body fat distribution in women compared to men[10,11]. Of the novel female (*DSTYK*
770 and *ANGPTL4*) and male (*UGGT2* and *MMP14*) specific signals, only *ANGPTL4* implicated fat distribution
771 related biology associated with both lipid biology and cardiovascular traits (**Box 1**). Sexual dimorphism in
772 fat distribution is apparent from childhood and throughout adult life[58-60], and at sexually dimorphic loci,
773 hormones with different levels in men and women may interact with genomic and epigenomic factors to
774 regulate gene activity, though this remains to be experimentally documented. Dissecting the underlying

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

778 Overall, we observe fewer significant associations between WHRadjBMI and coding variants on
779 the ExomeChip than Turcot *et al.* [25] examining the association of low frequency and rare coding variants
780 with BMI. In line with these observations, we identify fewer pathways and cross-trait associations. One
781 reason for fewer WHRadjBMI implicated variants and pathways may be smaller sample size ($N_{WHRadjBMI}$ =
782 476,546, N_{BMI} = 718,639), and thus, lower statistical power. Power, however, is likely not the only
783 contributing factor. For example, Turcot *et al.* [25] have comparative sample sizes between BMI and that of
784 Marouli *et al.*[22] studying height (N_{height} = 711,428). However, greater than seven times the number of
785 coding variants are identified for height than for BMI, indicating that perhaps a number of other factors,
786 including trait architecture, heritability (possibly overestimated in some phenotypes), and phenotype
787 precision, likely all contribute to our study's capacity to identify low frequency and rare variants with large
788 effects. Further, it is possible that the comparative lack of significant findings for WHRadjBMI and BMI
789 compared to height may be a result of higher selective pressure against genetic predisposition to
790 cardiometabolic phenotypes, such as BMI and WHR. As evolutionary theory predicts that harmful alleles
791 will be low frequency[61], we may need larger sample sizes to detect rare variants that have so far escaped
792 selective pressures. Lastly, the ExomeChip is limited by the variants that are present on the chip, which
793 was largely dictated by sequencing studies in European-ancestry populations and a MAF detection criteria
794 of ~0.012%. It is likely that through an increased sample size, use of chips designed to detect variation
795 across a range of continental ancestries, high quality, deep imputation with large reference samples (e.g.
796 HRC), and/or alternative study designs, future studies will detect additional variation from the entire allele
797 frequency spectrum that contributes to fat distribution phenotypes.

798 The collected genetic and epidemiologic evidence has now demonstrated that fat distribution (as
799 measured by increased WHRadjBMI) is correlated with increased risk of T2D and CVD, and that this
800 association is likely causal with potential mediation through blood pressure, triglyceride-rich lipoproteins,
801 glucose, and insulin[9]. This observation yields an immediate follow-up question: Which mechanisms
802 regulate depot-specific fat accumulation and are risks for disease, driven by increased visceral or
803 decreased subcutaneous adipose tissue mass (or both)? Pathway analysis identified several novel
804 pathways and gene sets related to metabolism and adipose regulation, bone growth and development
805 we also observed a possible role for adiponectin, a hormone which has been linked to "healthy" expansion
806 of adipose tissue and insulin sensitivity [62]. Similarly, expression/eQTL results support the function and
807 relevance of adipogenesis, adipocyte biology, and insulin signaling, supporting our previous findings for
808 WHRadjBMI[10]. We also provide evidence suggesting known biological functions and pathways
809 contributing to body fat distribution (e.g., diet-induced obesity, angiogenesis, bone growth and
810 morphology, and enhanced lipolysis).

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

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

829 A seminal finding from this study is the importance of lipid metabolism for body fat distribution.
830 In fact, pathway analyses that highlight enhanced lipolysis, cross-trait associations with circulating lipid
831 levels, existing biological evidence from the literature, and knockdown experiments in *Drosophila*
832 examining triglyceride storage point to novel candidate genes (*ANGPTL4*, *ACVR1C*, *DAGLB*, *MGA*, *RASIP1*,
833 and *IZUMO1*) and new candidates in known regions (*DNAH10*[10] and *MLXIPL*[14]) related to lipid biology and
834 its role in fat storage. Newly implicated genes of interest include *ACVR1C*, *MLXIPL*, and *ANGPTL4*, all of
835 which are involved in lipid homeostasis; all are excellent candidate genes for central adiposity. Carriers of
836 inactivating mutations in *ANGPTL4* (*Angiopoietin Like 4*), for example, display low triglyceride levels and
837 low risk of coronary artery disease[63]. *ACVR1C* encodes the activin receptor-like kinase 7 protein (ALK7), a
838 receptor for the transcription factor TGFB-1, well known for its central role in growth and development in
839 general[64-68], and adipocyte development in particular[68]. *ACVR1C* exhibits the highest expression in adipose
840 tissue, but is also highly expressed in the brain[69-71]. In mice, decreased activity of *ACVR1C* upregulates
841 PPARγ and C/EBPα pathways and increases lipolysis in adipocytes, thus decreasing weight and diabetes in
842 mice[69,72,73]. Such activity is suggestive of a role for ALK7 in adipose tissue signaling and therefore for
843 therapeutic targets for human obesity. *MLXIPL*, also important for lipid metabolism and postnatal cellular
844 growth, is a transcription factor which activates triglyceride synthesis genes in a glucose-dependent
845 manner[74,75]. The lead exome variant in this gene is highly conserved, most likely damaging, and is

846 associated with reduced *MLXIPL* expression in adipose tissue. Furthermore, in a recent longitudinal, *in*
847 *vitro* transcriptome analysis of adipogenesis in human adipose-derived stromal cells, gene expression of
848 *MLXIPL* was up-regulated during the maturation of adipocytes, suggesting a critical role in the regulation
849 of adipocyte size and accumulation[76]. However, given our observations on cross-trait associations with
850 variants in *MLXIPL* and diabetes-related traits, development of therapeutic targets must be approached
851 cautiously.

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

859

860

861 **ACKNOWLEDGEMENTS**

862 A full list of acknowledgements is provided in the **Supplementary Table 17**. This study was completed as
863 part of the Genetic Investigation of ANtropometric Traits (GIANT) Consortium. This research has been
864 conducted using the UK Biobank resource. Funding for this project was provided by Aase and Ejner
865 Danielsens Foundation, Academy of Finland (102318; 123885; 117844; 40758; 211497; 118590; 139635;
866 129293; 286284; 134309; 126925; 121584; 124282; 129378; 117787; 41071; 137544; 272741), Action on
867 Hearing Loss (G51), ALK-Abelló A/S (Hørsholm-Denmark), American Heart Association (13EIA14220013;
868 13GRNT16490017; 13POST16500011), American Recovery and Reinvestment Act of 2009 (ARRA)

869 Supplement (EY014684-03S1; -04S1), Amgen, André and France Desmarais Montreal Heart Institute (MHI)
870 Foundation, AstraZeneca, Augustinus Foundation, Australian Government and Government of Western
871 Australia, Australian Research Council Future Fellowship, Becket Foundation, Benzon Foundation, Bernard
872 Wolfe Health Neuroscience Endowment, British Heart Foundation (CH/03/001; RG/14/5/30893;
873 RG/200004; SP/04/002; SP/09/002), BiomarCaRE (278913), Bundesministerium für Bildung und
874 Forschung (Federal Ministry of Education and Research-Germany; German Center for Diabetes Research
875 (DZD); 01ER1206; 01ER1507; 01ER1206; 01ER1507; FKZ: 01EO1501 (AD2-060E); 01ZZ9603; 01ZZ0103;
876 01ZZ0403; 03IS2061A; 03Z1CN22; FKZ 01GI1128), Boehringer Ingelheim Foundation, Boston University
877 School of Medicine, Canada Research Chair program, Canadian Cancer Society Research Institute,
878 Canadian Institutes of Health Research (MOP-82893), Cancer Research UK (C864/A14136; A490/A10124;
879 C8197/A16565), Cebu Longitudinal Health and Nutrition Survey (CLHNS) pilot funds (RR020649;
880 ES010126; DK056350), Center for Non-Communicable Diseases (Pakistan), Central Society for Clinical
881 Research, Centre National de Génotypage (Paris-France), CHDI Foundation (Princeton-USA), Chief
882 Scientist Office of the Scottish Government Health Directorate (CZD/16/6), City of Kuopio and Social
883 Insurance Institution of Finland (4/26/2010), Clarendon Scholarship, Commission of the European
884 Communities; Directorate C-Public Health (2004310), Copenhagen County, County Council of Dalarna,
885 Curtin University of Technology, Dalarna University, Danish Centre for Evaluation and Health Technology
886 Assessment, Danish Council for Independent Research, Danish Diabetes Academy, Danish Heart
887 Foundation, Danish Medical Research Council-Danish Agency for Science Technology and Innovation,
888 Danish Medical Research Council, Danish Pharmaceutical Association, Danish Research Council for
889 Independent Research, Dekker scholarship (2014T001), Dentistry and Health Sciences, Department of
890 Internal Medicine at the University of Michigan, Diabetes Care System West-Friesland, Diabetes Heart
891 Study (R01 HL6734; R01 HL092301; R01 NS058700), Doris Duke Charitable Foundation Clinical Scientist
892 Development Award (2014105), Doris Duke Medical Foundation, Dr. Robert Pfleger Stiftung, Dutch Cancer

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

917 Epidemiology of Colorectal Cancer Consortium (NCI CA137088), German Cancer Aid (70-2488-Ha I),
918 German Diabetes Association, German Research Foundation (CRC 1052 C01; B01; B03), Health and
919 Retirement Study (R03 AG046398), Health Insurance Foundation (2010 B 131), Health Ministry of
920 Lombardia Region (Italy), Helmholtz Zentrum München – German Research Center for Environmental
921 Health, Helse Vest, Home Office (780-TETRA), Hospital Districts of Pirkanmaa; Southern Ostrobothnia;
922 North Ostrobothnia; Central Finland and Northern Savo, Ib Henriksen Foundation, Imperial College
923 Biomedical Research Centre, Imperial College Healthcare NHS Trust, Institute of Cancer Research and The
924 Everyman Campaign, Interuniversity Cardiology Institute of the Netherlands (09.001), Intramural
925 Research Program of the National Institute on Aging, Italian Ministry of Health (GR-2011-02349604), Johns
926 Hopkins University School of Medicine (HHSN268200900041C), Juho Vainio Foundation, Kaiser
927 Foundation Research Institute (HHSN268201300029C), KfH Stiftung Präventivmedizin e.V., KG Jebsen
928 Foundation, Knut and Alice Wallenberg Foundation (Wallenberg Academy Fellow), Knut och Alice
929 Wallenberg Foundation (2013.0126), Kuopio Tampere and Turku University Hospital Medical Funds
930 (X51001), Kuopio University Hospital, Leenaards Foundation, Leiden University Medical Center, Li Ka Shing
931 Foundation (CML), Ludwig-Maximilians-Universität, Lund University, Lundbeck Foundation,
932 Lundbeckfonden, Marianne and Marcus Wallenberg Foundation, Max Planck Society, Medical Research
933 Council-UK (G0601966; G0700931; G0000934; MR/L01632X/1; MC_UU_12015/1; MC_PC_13048;
934 G9521010D; G1000143; MC_UU_12013/1-9; MC_UU_12015/1; MC_PC_13046; MC_U106179471;
935 G0800270, MR/L01341X/1), MEKOS Laboratories (Denmark), Merck & Co Inc., MESA Family (R01-HL-
936 071205; R01-HL-071051; R01-HL-071250; R01-HL-071251; R01-HL-071252; R01-HL-071258; R01-HL-
937 071259; UL1-RR-025005), Ministry for Health Welfare and Sports (the Netherlands), Ministry of Cultural
938 Affairs (Germany), Ministry of Education and Culture of Finland (627;2004-2011), Ministry of Education
939 Culture and Science (the Netherlands), Ministry of Science and Technology (Taiwan) (MOST 104-2314-B-
940 075A-006 -MY3), Ministry of Social Affairs and Health in Finland, Montreal Heart Institute Foundation,

941 MRC-PHE Centre for Environment and Health, Multi-Ethnic Study of Atherosclerosis (MESA) (N01-HC-
942 95159; N01-HC-95160; N01-HC-95161; N01-HC-95162; N01-HC-95163; N01-HC-95164; N01-HC-95165;
943 N01-HC-95166; N01-HC-95167; N01-HC-95168; N01-HC-95169), Munich Center of Health Sciences (MC-
944 Health), Municipality of Rotterdam (the Netherlands) Murdoch University, National Basic Research
945 Program of China (973 Program 2012CB524900), National Cancer Institute (CA047988; UM1CA182913),
946 National Cancer Research Institute UK, National Cancer Research Network UK, National Center for
947 Advancing Translational Sciences (UL1TR001881), National Center for Research Resources (UL1-TR-
948 000040 and UL1-RR-025005), National Eye Institute of the National Institutes of Health (EY014684, EY-
949 017337), National Health and Medical Research Council of Australia (403981; 1021105; 572613), National
950 Heart Lung and Blood Institute (HHSN268201100010C; HHSN268201100011C; and HHSN268201100012C;
951 HHSN268201100005; HHSN268201100006C; HHSN268201100007C; HHSN268201100008C;
952 HHSN268201100009; HHSN268201100037C; HHSN268201300046C; HHSN268201300047C;
953 HHSN268201300048C; HHSN268201300049C; HHSN268201300050C; HL043851; HL080467; HL094535;
954 HL109946; HHSN268201300025C; HHSN268201300026C; HL119443; HL054464; HL054457; HL054481;
955 HL087660; HL086694; HL060944; HL061019; HL060919; HL060944; HL061019; N02-HL-6-4278; R21
956 HL121422-02; R21 HL121422-02; R01 DK089256-05), National Human Genome Research Institute
957 (HG007112), National Institute for Health Research BioResource Clinical Research Facility and Biomedical
958 Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London, National
959 Institute for Health Research Comprehensive Biomedical Research Centre Imperial College Healthcare
960 NHS Trust, National Institute for Health Research (NIHR) (RP-PG-0407-10371), National Institute of
961 Diabetes and Digestive and Kidney Disease (DK063491; DK097524; DK085175; DK087914;
962 1R01DK8925601; 1R01DK106236-01A1), National Institute of Health Research Senior Investigator,
963 National Institute on Aging (NIA U01AG009740; RC2 AG036495; RC4 AG039029), National Institute on
964 Minority Health and Health Disparities, National Institutes of Health (NIH) (1R01HG008983-01;

965 1R21DA040177-01; 1RO1HL092577; R01HL128914; K24HL105780; K01HL116770; U01 HL072515-06; U01
966 HL84756; U01HL105198; U01 GM074518; R01 DK089256-05; R01DK075787; R25 CA94880; P30
967 CA008748; DK078150; TW005596; HL085144; TW008288; R01-HL093029; U01- HG004729; R01-
968 DK089256; 1R01DK101855-01; 1K99HL130580; T32-GM067553; U01-DK105561; R01-HL-117078; R01-DK-
969 089256; UO1HG008657; UO1HG06375; UO1AG006781; DK064265; R01DK106621-01; K23HL114724;
970 NS33335; HL57818; R01-DK089256; 2R01HD057194; U01HG007416; R01DK101855, R01DK075787, T32
971 GM096911-05; K01 DK107836; R01DK075787; UO1 AG 06781; U01-HG005152, 1F31HG009850-01),
972 National Key R&D Plan of China (2016YFC1304903), Key Project of the Chinese Academy of Sciences
973 (ZDBS-SSW-DQC-02, ZDRW-ZS-2016-8-1, KJZD-EW-L14-2-2), National Natural Science Foundation of China
974 (81471013; 30930081; 81170734; 81321062; 81471013), National NIHR Bioresource, National Science
975 Council (Taiwan) (NSC 102-2314-B-075A-002), Netherlands CardioVascular Research Initiative
976 (CVON2011-19), Netherlands Heart Foundation, Netherlands Organisation for Health Research and
977 Development (ZonMW) (113102006), Netherlands Organisation for Scientific Research (NWO)-sponsored
978 Netherlands Consortium for Healthy Aging (050-060-810), Netherlands Organization for Scientific
979 Research (184021007), NHMRC Practitioner Fellowship (APP1103329), NIH through the American
980 Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419), NIHR Biomedical Research Centre at
981 The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, NIHR Cambridge
982 Biomedical Research Centre, NIHR Cambridge Biomedical Research Centre, NIHR Health Protection
983 Research Unit on Health Impact of Environmental Hazards (HPRU-2012-10141), NIHR Leicester
984 Cardiovascular Biomedical Research Unit, NIHR Official Development Assistance (ODA, award 16/136/68),
985 NIHR Oxford Biomedical Research Centre, the European Union FP7 (EpiMigrant, 279143) and H2020
986 programs (iHealth-T2D; 643774), NIHR Senior Investigator, Nordic Centre of Excellence on Systems Biology
987 in Controlled Dietary Interventions and Cohort Studies (SYSDIET) (070014), Northwestern University
988 (HHSN268201300027C), Norwegian Diabetes Association, Novartis, Novo Nordisk Foundation, Nuffield

989 Department of Clinical Medicine Award, Orchid Cancer Appeal, Oxford Biomedical Research Centre, Paavo
990 Nurmi Foundation, Päivikki and Sakari Sohlberg Foundation, Pawsey Supercomputing Centre (funded by
991 Australian Government and Government of Western Australia), Peninsula Research Bank-NIHR Exeter
992 Clinical Research Facility, Pfizer, Prostate Cancer Research Foundation, Prostate Research Campaign UK
993 (now Prostate Action), Public Health England, QIMR Berghofer, Raine Medical Research Foundation,
994 Regione FVG (L.26.2008), Republic of Croatia Ministry of Science, Research Centre for Prevention and
995 Health-the Capital Region of Denmark, Research Council of Norway, Research Institute for Diseases in the
996 Elderly (RIDE), Research into Ageing, Robert Dawson Evans Endowment of the Department of Medicine
997 at Boston University School of Medicine and Boston Medical Center, Science Live/Science Center NEMO,
998 Scottish Funding Council (HR03006), Sigrid Juselius Foundation, Social Insurance Institution of Finland,
999 Singapore Ministry of Health’s National Medical Research Council (NMRC/STaR/0028/2017), Social
1000 Ministry of the Federal State of Mecklenburg-West Pomerania, State of Bavaria-Germany, State of
1001 Washington Life Sciences Discovery Award (265508) to the Northwest Institute of Genetic Medicine,
1002 Stroke Association, Swedish Diabetes Foundation (2013-024), Swedish Heart-Lung Foundation (20120197;
1003 20120197; 20140422), Swedish Research Council (2012-1397), Swedish Research Council Strategic
1004 Research Network Epidemiology for Health, Swiss National Science Foundation (31003A-143914),
1005 SystemsX.ch (51RTP0_151019), Taichung Veterans General Hospital (Taiwan) (TCVGH-1047319D; TCVGH-
1006 1047311C), Tampere Tuberculosis Foundation, TEKES Grants (70103/06; 40058/07), The Telethon Kids
1007 Institute, Timber Merchant Vilhelm Bangs Foundation, UCL Hospitals NIHR Biomedical Research Centre,
1008 UK Department of Health, Université de Montréal Beaulieu-Saucier Chair in Pharmacogenomics,
1009 University Hospital Regensburg, University of Bergen, University of Cambridge, University of Michigan
1010 Biological Sciences Scholars Program, University of Michigan Internal Medicine Department Division of
1011 Gastroenterology, University of Minnesota (HHSN268201300028C), University of Notre Dame (Australia),
1012 University of Queensland, University of Western Australia (UWA), Uppsala Multidisciplinary Center for

1013 Advanced Computational Science (b2011036), Uppsala University, US Department of Health and Human
1014 Services (HHSN268201100046C; HHSN268201100001C; HHSN268201100002C; HHSN268201100003C;
1015 HHSN268201100004C; HHSN271201100004C), UWA Faculty of Medicine, Velux Foundation, Wellcome
1016 Trust (083948/B/07/Z; 084723/Z/08/Z; 090532; 098381; 098497/Z/12/Z; WT098051; 068545/Z/02;
1017 WT064890; WT086596; WT098017; WT090532; WT098051; WT098017; WT098381; WT098395; 083948;
1018 085475), Western Australian DNA Bank (National Health and Medical Research Council of Australia
1019 National Enabling Facility), Women and Infant's Research Foundation, Yrjš Jahnsson Foundation (56358)
1020

1021 **AUTHORSHIP CONTRIBUTIONS**

1022 Writing Group: LAC, RSF, TMF, MG, HMH, JNH, AEJ, TK, ZK, CML, RJFL, YL, KEN, VT, KLY; Data preparation
1023 group: TA, IBB, TE, SF, MG, HMH, AEJ, TK, DJL, KSL, AEL, RJFL, YL, EM, NGDM, MCMG, PM, MCYN, MAR,
1024 SS, CS, KS, VT, SV, SMW, TWW, KLY, XZ; WHR meta-analyses: PLA, HMH, AEJ, TK, MG, CML, RJFL, KEN, VT,
1025 KLY; Pleiotropy working group: GA, MB, JPC, PD, FD, JCF, HMH, SK, HK, HMH, AEJ, CML, DJL, RJFL, AM, EM,
1026 GM, MIM, PBM, GMP, JRBP, KSR, XS, SW, JW, CJW; Phenome-wide association studies: LB, JCD, TLE, AG,
1027 AM, MIM; Gene-set enrichment analyses: SB, RSF, JNH, ZK, DL, THP; eQTL analyses: CKR, YL, KLM;
1028 Monogenic and syndromic gene enrichment analyses: HMH, AKM; Fly Obesity Screen: AL, JAP; Overseeing
1029 of contributing studies: (1958 Birth Cohort) PD; (Airwave) PE; (AMC PAS) GKH; (Amish) JRO'C; (ARIC) EB;
1030 (ARIC, Add Health) KEN; (BRAVE) EDA, RC; (BRIGHT) PBM; (CARDIA) MF, PJS; (Cebu Longitudinal Health
1031 and Nutrition Survey) KLM; (CHD Exome + Consortium) ASB, JMMH, DFR, JD; (CHES) RV; (Clear/eMERGE
1032 (Seattle)) GPJ; (CROATIA_Korcula) VV, OP, IR; (deCODE) KS, UT; (DHS) DWB; (DIACORE) CAB; (DPS) JT, JL,
1033 MU; (DRSEXTRA) TAL, RR; (EFSOCH) ATH, TMF; (EGCUT) TE; (eMERGE (Seattle)) EBL; (EPIC-Potsdam) MBS,
1034 HB; (EpiHealth) EI, PWF; (EXTEND) ATH, TMF; (Family Heart Study) IBB; (Fenland, EPIC) RAS; (Fenland,
1035 EPIC, InterAct) NJW, CL; (FINRISK) SM; (FINRISK 2007 (T2D)) PJ, VS; (Framingham Heart Study) LAC;
1036 (FUSION) MB, FSC; (FVG) PG; (Generation Scotland) CH, BHS; (Genetic Epidemiology Network of

1037 Arteriopathy (GENOA)) SLRK; (GRAPHIC) NJS; (GSK-STABILITY) DMW, LW, HDW; (Health) AL; (HELIC
1038 MANOLIS) EZ, GD; (HELIC Pomak) EZ, GD; (HUNT-MI) KH, CJW; (Inter99) TH, TJ; (IRASFS) LEW, EKS; (Jackson
1039 Heart Study (JHS)) JGW; (KORA S4) KS, IMH; (Leipzig-Adults) MB, PK; (LOLIPOP-Exome) JCC, JSK; (LOLIPOP-
1040 OmniEE) JCC, JSK; (MESA) JIR, XG; (METSIM) JK, ML; (MONICA-Brianza) GC; (Montreal Heart Institute
1041 Biobank (MHIBB)) MPD, GL, SdD, JCT; (MORGAM Central Laboratory) MP; (MORGAM Data Centre) KK;
1042 (OBB) FK; (PCOS) APM, CML; (PIVUS) CML, LL; (PRIME - Belfast) FK; (PRIME - Lille) PA; (PRIME - Strasbourg)
1043 MM; (PRIME - Toulouse) JF; (PROMIS) DS; (QC) MAR; (RISC) BB, EF, MW; (Rotterdam Study I) AGU, MAI;
1044 (SEARCH) AMD; (SHIP/SHIP-Trend) MD; (SIBS) DFE; (SOLID TIMI-52) DMW; (SORBS) APM, MS, AT; (The
1045 Mount Sinai BioMe Biobank) EPB, RJFL; (The NEO Study) DOMK; (The NHAPC study, The GBTDS study) XL;
1046 (The Western Australian Pregnancy Cohort (Raine) Study) CEP, SM; (TwinsUK) TDS; (ULSAM) APM; (Vejle
1047 Biobank) IB, CC, OP; (WGHS) DIC, PMR; (Women's Health Initiative) PLA; (WTCCC-UKT2D) MIM, KRO; (YFS)
1048 TL, OTRa; Genotyping of contributing studies: (1958 Birth Cohort) KES; (Airwave) EE, MPSL; (AMC PAS) SS;
1049 (Amish) LMYA, JAP; (ARIC) EWD, MG; (BBMRI-NL) SHV, LB, CMvD, PIWdB; (BRAVE) EDA; (Cambridge
1050 Cancer Studies) JGD; (CARDIA) MF; (CHD Exome + Consortium) ASB, JMMH, DFR, JD, RY(Clear/eMERGE
1051 (Seattle)) GPJ; (CROATIA_Korcula) VV; (DIACORE) CAB, MG; (DPS) AUJ, JL; (DRSEXTRA) PK; (EGCUT) TE;
1052 (EPIC-Potsdam) MBS, KM; (EpiHealth) EI, PWF; (Family Heart Study) KDT; (Fenland, EPIC) RAS; (Fenland,
1053 EPIC, InterAct) NJW, CL; (FUSION) NN; (FVG) IG, AM; (Generation Scotland) CH; (Genetic Epidemiology
1054 Network of Arteriopathy (GENOA)) SLRK, JAS; (GRAPHIC) NJS; (GSK-STABILITY) DMW; (Health) JBJ; (HELIC
1055 MANOLIS) LS; (HELIC Pomak) LS; (Inter99) TH, NG; (KORA) MMN; (KORA S4) KS, HG; (Leipzig-Adults) AM;
1056 (LOLIPOP-Exome) JCC, JSK; (LOLIPOP-OmniEE) JCC, JSK; (MESA) JIR, YDIC, KDT; (METSIM) JK, ML; (Montreal
1057 Heart Institute Biobank (MHIBB)) MPD; (OBB) FK; (PCOS) APM; (PIVUS) CML; (Rotterdam Study I) AGU,
1058 CMG, FR; (SDC) JMJ, HV; (SEARCH) AIMD; (SOLID TIMI-52) DMW; (SORBS) APM; (The Mount Sinai BioMe
1059 Biobank) EPB, RJFL, YL, CS; (The NEO Study) RLG; (The NHAPC study, The GBTDS study) XL, HL, YH; (The
1060 Western Australian Pregnancy Cohort (Raine) Study) CEP, SM; (TUDR) ZA; (TwinsUK) APM; (ULSAM) APM;

1061 (WGHS) DIC, AYC; (Women's Health Initiative) APR; (WTCCC-UKT2D) MIM; (YFS) TL, LPL; Phenotyping of
1062 contributing studies: (Airwave) EE; (AMC PAS) SS; (Amish) LM YA; (ARIC) EWD; (ARIC, Add Health) KEN;
1063 (BBMRI-NL) SHV; (BRAVE) EDA; (BRIGHT) MJC; (CARL) AR, GG; (Cebu Longitudinal Health and Nutrition
1064 Survey) NRL; (CHES) RV, MT; (Clear/eMERGE (Seattle)) GPJ, AAB; (CROATIA_Korcula) OP, IR; (DIACORE)
1065 CAB, BKK; (DPS) AUJ, JL; (EFSOCH) ATH; (EGCUT) EM; (EPIC-Potsdam) HB; (EpiHealth) EI; (EXTEND) ATH;
1066 (Family Heart Study) MFF; (Fenland, EPIC, InterAct) NJW; (FIN-D2D 2007) LM, MV; (FINRISK) SM; (FINRISK
1067 2007 (T2D)) PJ, HS; (Framingham Heart Study) CSF; (Generation Scotland) CH, BHS; (Genetic Epidemiology
1068 Network of Arteriopathy (GENOA)) SLRK, JAS; (GRAPHIC) NJS; (GSK-STABILITY) LW, HDW; (Health) AL, BHT;
1069 (HELIC MANOLIS) LS, AEF, ET; (HELIC Pomak) LS, AEF, MK; (HUNT-MI) KH, OH; (Inter99) TJ, NG; (IRASFS)
1070 LEW, BK; (KORA) MMN; (LASA (BBMRI-NL)) KMAS; (Leipzig-Adults) MB, PK; (LOLIPOP-Exome) JCC, JSK;
1071 (LOLIPOP-OmniEE) JCC, JSK; (MESA) MA; (Montreal Heart Institute Biobank (MHIBB)) GL, KSL, VT;
1072 (MORGAM Data Centre) KK; (OBB) FK, MN; (PCOS) CML; (PIVUS) LL; (PRIME - Belfast) FK; (PRIME - Lille)
1073 PA; (PRIME - Strasbourg) MM; (PRIME - Toulouse) JF; (RISC) BB, EF; (Rotterdam Study I) MAI, CMGFR, MCZ;
1074 (SHIP/SHIP-Trend) NF; (SORBS) MS, AT; (The Mount Sinai BioMe Biobank) EPB, YL, CS; (The NEO Study)
1075 RdM; (The NHAPC study, The GBTDS study) XL, HL, LS, FW; (The Western Australian Pregnancy Cohort
1076 (Raine) Study) CEP; (TUDR) YJH, WJL; (TwinsUK) TDS, KSS; (ULSAM) VG; (WGHS) DIC, PMR; (Women's
1077 Health Initiative) APR; (WTCCC-UKT2D) MIM, KRO; (YFS) TL, OTR; Data analysis of contributing studies:
1078 (1958 Birth Cohort) KES, IN; (Airwave) EE, MPSL; (AMC PAS) SS; (Amish) JRO'C, LMYA, JAP; (ARIC, Add
1079 Health) KEN, KLY, MG; (BBMRI-NL) LB; (BRAVE) RC, DSA; (BRIGHT) HRW; (Cambridge Cancer Studies) JGD,
1080 AP, DJT; (CARDIA) MF, LAL; (CARL) AR, DV; (Cebu Longitudinal Health and Nutrition Survey) YW; (CHD
1081 Exome + Consortium) ASB, JMMH, DFR, RY, PS; (CHES) YJ; (CROATIA_Korcula) VV; (deCODE) VS, GT; (DHS)
1082 AJC, PM, MCYN; (DIACORE) CAB, MG; (EFSOCH) HY; (EGCUT) TE, RM; (eMERGE (Seattle)) DSC; (ENDO) TK;
1083 (EPIC) JHZ; (EPIC-Potsdam) KM; (EpiHealth) SG; (EXTEND) HY; (Family Heart Study) MFF; (Fenland) JaL;
1084 (Fenland, EPIC) RAS; (Fenland, InterAct) SMW; (Finrisk Extremes and QC) SV; (Framingham Heart Study)

1085 CTL, NLHC; (FVG) IG; (Generation Scotland) CH, JM; (Genetic Epidemiology Network of Arteriopathy
1086 (GENOA)) LFB; (GIANT-Analyst) AEJ; (GRAPHIC) NJS, NGDM, CPN; (GSK-STABILITY) DMW, AS; (Health) JBJ;
1087 (HELIC MANOLIS) LS; (HELIC Pomak) LS; (HUNT-MI) WZ; (Inter99) NG; (IRASFS) BK; (Jackson Heart Study
1088 (JHS)) LAL, JL; (KORA S4) TWW; (LASA (BBMRI-NL)) KMAS; (Leipzig-Adults) AM; (LOLIPOP-Exome) JCC, JSK,
1089 WZ; (LOLIPOP-OmniEE) JCC, JSK, WZ; (MESA) JIR, XG, JY; (METSIM) XS; (Montreal Heart Institute Biobank
1090 (MHIBB)) JCT, GL, KSL, VT; (OBB) AM; (PCOS) APM, TK; (PIVUS) NR; (PROMIS) AR, WZ; (QC GoT2D/T2D-
1091 GENES (FUSION, METSIM, etc)) AEL; (RISC) HY; (Rotterdam Study I) CMG, FR; (SHIP/SHIP-Trend) AT; (SOLID
1092 TIMI-52) DMW, AS; (SORBS) APM; (The Mount Sinai BioMe Biobank) YL, CS; (The NEO Study) RLG; (The
1093 NHAPC study, The GBTDS study) XL, HL, YH; (The Western Australian Pregnancy Cohort (Raine) Study)
1094 CAW; (UK Biobank) ARW; (ULSAM) APM, AM; (WGHS) DIC, AYC; (Women's Health Initiative) PLA, JH;
1095 (WTCCC-UKT2D) WG; (YFS) LPL.

1096

1097 # METHODS

1098 ## Studies

1099 Stage 1 consisted of 74 studies (12 case/control studies, 59 population-based studies, and five
1100 family studies) comprising 344,369 adult individuals of the following ancestries: 1) European descent (N=
1101 288,492), 2) African (N= 15,687), 3) South Asian (N= 29,315), 4) East Asian (N=6,800), and 5) Hispanic
1102 (N=4,075). Stage 1 meta-analyses were carried out in each ancestry separately and in the all ancestries
1103 group, for both sex-combined and sex-specific analyses. Follow-up analyses were undertaken in 132,177
1104 individuals of European ancestry from the deCODE anthropometric study and UK Biobank (**Supplementary
1105 Tables 1-3**). Conditional analyses were performed in the all ancestries and European descent groups.
1106 Informed consent was obtained for participants by the parent study and protocols approved by each
1107 study's institutional review boards.

1108 **Phenotypes**

1109 For each study, WHR (waist circumference divided by hip circumference) was corrected for age,
1110 BMI, and the genomic principal components (derived from GWAS data, the variants with MAF >1% on the
1111 ExomeChip, and ancestry informative markers available on the ExomeChip), as well as any additional
1112 study-specific covariates (e.g. recruiting center), in a linear regression model. For studies with non-related
1113 individuals, residuals were calculated separately by sex, whereas for family-based studies sex was included
1114 as a covariate in models with both men and women. Additionally, residuals for case/control studies were
1115 calculated separately. Finally, residuals were inverse normal transformed and used as the outcome in
1116 association analyses. Phenotype descriptives by study are shown in **Supplementary Table 3**.

1117 **Genotypes and QC**

1118 The majority of studies followed a standardized protocol and performed genotype calling using
1119 the algorithms indicated in **Supplementary Table 2**, which typically included zCall[3]. For 10 studies
1120 participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium,
1121 the raw intensity data for the samples from seven genotyping centers were assembled into a single project
1122 for joint calling[4]. Study-specific quality control (QC) measures of the genotyped variants were
1123 implemented before association analysis (**Supplementary Tables 1-2**). Furthermore, to assess the
1124 possibility that any significant associations with rare and low-frequency variants could be due to allele
1125 calling in the smaller studies, we performed a sensitivity meta-analysis including all large studies (>5,000
1126 participants) and compared to all studies. We found very high concordance for effect sizes, suggesting
1127 that smaller studies do not bias our results (**Supplementary Fig. 24**).

1128 **Study-level statistical analyses**

1129 Individual cohorts were analyzed for each ancestry separately, in sex-combined and sex-specific
1130 groups, with either RAREMETALWORKER (http://genome.sph.umich.edu/wiki/RAREMETALWORKER) or

1131 RVTESTs (http://zhanxw.github.io/rvtests/), to associate inverse normal transformed WHRadjBMI with
1132 genotype accounting for cryptic relatedness (kinship matrix) in a linear mixed model. These software
1133 programs are designed to perform score-statistic based rare-variant association analysis, can
1134 accommodate both unrelated and related individuals, and provide single-variant results and variance-
1135 covariance matrices. The covariance matrix captures linkage disequilibrium (LD) relationships between
1136 markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses[77,78]. Single-
1137 variant analyses were performed for both additive and recessive models.

1138 **Centralized quality-control**

1139 Individual cohorts identified ancestry population outliers based on 1000 Genome Project phase 1
1140 ancestry reference populations. A centralized quality-control procedure implemented in EasyQC[79] was
1141 applied to individual cohort association summary statistics to identify cohort-specific problems: (1)
1142 assessment of possible errors in phenotype residual transformation; (2) comparison of allele frequency
1143 alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues,
1144 and (3) examination of quantile-quantile (QQ) plots per study to identify any inflation arising from
1145 population stratification, cryptic relatedness and genotype biases.

1146 **Meta-analyses**

1147 Meta-analyses were carried out in parallel by two different analysts at two sites using
1148 RAREMETAL[77]. During the meta-analyses, we excluded variants if they had call rate <95%, Hardy-Weinberg
1149 equilibrium P-value $<1 \times 10^{-7}$, or large allele frequency deviations from reference populations (>0.6 for all
1150 ancestries analyses and >0.3 for ancestry-specific population analyses). We also excluded from
1151 downstream analyses markers not present on the Illumina ExomeChip array 1.0, variants on the Y-
1152 chromosome or the mitochondrial genome, indels, multiallelic variants, and problematic variants based
1153 on the Blat-based sequence alignment analyses. Significance for single-variant analyses was defined at an

1154 array-wide level ($P<2x10^{-7}$). For all suggestive significant variants from Stage 1, we tested for significant
1155 sex differences. We calculated Psexhet for each SNP, testing for difference between women-specific and
1156 men-specific beta estimates and standard errors using EasyStrata[11,80]. Each SNP that reached
1157 $P_{sexhet}<0.05$/# of variants tested (70 variants brought forward from Stage 1, $P_{sexhet}<7.14x10^{-4}$) was
1158 considered significant. Additionally, while each individual study was asked to perform association analyses
1159 stratified by race/ethnicity, and adjust for population stratification, all study-specific summary statistics
1160 were meta-analyzed together for our all ancestry meta-analyses. To investigate potential heterogeneity
1161 across ancestries, we did examine ancestry-specific meta-analysis results for our top 70 variants from
1162 stage 1, and found no evidence of significant across-ancestry heterogeneity observed for any of our top
1163 variants (I^2 values noted in **Supplementary Data 1-3**).

1164 For the gene-based analyses, we applied two sets of criteria to select variants with a MAF<5%
1165 within each ancestry based on coding variant annotation from five prediction algorithms (PolyPhen2,
1166 HumDiv and HumVar, LRT, MutationTaster, and SIFT)[80,81]. Our broad gene-based tests included nonsense,
1167 stop-loss, splice site, and missense variants annotated as damaging by at least one algorithm mentioned
1168 above. Our strict gene-based tests included only nonsense, stop-loss, splice site, and missense variants
1169 annotated as damaging by all five algorithms. These analyses were performed using the sequence kernel
1170 association test (SKAT) and variable threshold (VT) methods. Statistical significance for gene-based tests
1171 was set at a Bonferroni-corrected threshold of $P<2.5x10^{-6}$ (0.05/~20,000 genes). All gene-based tests were
1172 performed in RAREMETAL[77].

1173 **Genomic inflation**

1174 We observed a marked genomic inflation of the test statistics even after controlling for population
1175 stratification (linear mixed model) arising mainly from common markers; λ_{GC} in the primary meta-analysis
1176 (combined ancestries and combined sexes) was 1.06 and 1.37 for all and only common coding and splice
1177 site markers considered herein, respectively (**Supplementary Figures 3, 7** and **13**, **Supplementary Table**

1178 **16**). Such inflation is expected for a highly polygenic trait like WHRadjBMI, for studies using a non-random

1179 set of variants across the genome, and is consistent with our very large sample size[79,82,83].

1180 **Conditional analyses**

1181 The RAREMETAL R-package[77] was used to identify independent WHRadjBMI association signals

1182 across all ancestries and European meta-analysis results. RAREMETAL performs conditional analyses by

1183 using covariance matrices to distinguish true signals from the shadows of adjacent significant variants in

1184 LD. First, we identified the lead variants ($P<2x10^{-7}$) based on a 1Mb window centered on the most

1185 significantly associated variant. We then conditioned on the lead variants in RAREMETAL and kept new

1186 lead signals at $P<2x10^{-7}$ for conditioning in a second round of analysis. The process was repeated until no

1187 additional signal emerged below the pre-specified P-value threshold ($P<2x10^{-7}$).

1188 To test if the associations detected were independent of the previously published WHRadjBMI

1189 variants [10,14,16], we performed conditional analyses in the stage 1 discovery set if the GWAS variant or its

1190 proxy ($r^2\geq0.8$) was present on the ExomeChip using RAREMETAL[77]. All variants identified in our meta-

1191 analysis and the previously published variants were also present in the UK Biobank dataset[84]. This dataset

1192 was used as a replacement dataset if a good proxy was not present on the ExomeChip as well as a

1193 replication dataset for the variants present on the ExomeChip. All conditional analyses in the UK Biobank

1194 dataset were performed using SNPTEST[85-87]. The conditional analyses were carried out reciprocally,

1195 conditioning on the ExomeChip variant and then the previously published variant. An association was

1196 considered independent of the previously published association if there was a statistically significant

1197 association detected prior to the conditional analysis ($P<2x10^{-7}$) with both the exome chip variant and the

1198 previously published variant, and the observed association with both or either of the variants disappeared

1199 upon conditional analysis ($P>0.05$). A conditional p-value between $9x10^{-6}$ and 0.05 was considered

1200 inconclusive. However, a conditional p-value $<9x10^{-6}$ was also considered suggestive.

1201

1202 **Stage 2 meta-analyses**

1203 In our Stage 2, we sought to validate a total of 70 variants from Stage 1 that met $P<2x10^{-6}$ in two
1204 independent studies, the UK Biobank (Release 1[84]) and Iceland (deCODE), comprising 119,572 and 12,605
1205 individuals, respectively (Supplementary Tables 1-3). The same QC and analytical methodology were used
1206 for these studies. Genotyping, study descriptions and phenotype descriptives are provided in
1207 **Supplementary Tables 1-3**. For the combined analysis of Stage 1 plus 2, we used the inverse-variance
1208 weighted fixed effects meta-analysis method. Significant associations were defined as those nominally
1209 significant (P<0.05) in the Stage 2 study and for the combined meta-analysis (Stage 1 plus Stage 2)
1210 significance was set at $P<2x10^{-7}$ (0.05/~250,000 variants).

1211 **Pathway enrichment analyses: EC-DEPICT**

1212 We adapted DEPICT, a gene set enrichment analysis method for GWAS data, for use with the
1213 ExomeChip ('EC-DEPICT'); this method is also described in a companion manuscript[22]. DEPICT's primary
1214 innovation is the use of "reconstituted" gene sets, where many different types of gene sets (e.g. canonical
1215 pathways, protein-protein interaction networks, and mouse phenotypes) were extended through the use
1216 of large-scale microarray data (see Pers et al.[21] for details). EC-DEPICT computes p-values based on
1217 Swedish ExomeChip data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania (ANDIS), and Scania
1218 Diabetes Registry (SDR) cohorts, N=11,899) and, unlike DEPICT, takes as input only the genes directly
1219 containing the significant (coding) variants rather than all genes within a specified amount of linkage
1220 disequilibrium (see **Supplementary Note 2**).

1221 Two analyses were performed for WHRadjBMI ExomeChip: one with all variants $p<5x10^{-4}$ (49
1222 significant gene sets in 25 meta-gene sets, FDR <0.05) and one with all variants > 1 Mb from known GWAS
1223 loci [10] (26 significant gene sets in 13 meta-gene sets, FDR <0.05). Affinity propagation clustering[88] was

1224 used to group highly correlated gene sets into "meta-gene sets"; for each meta-gene set, the member
1225 gene set with the best p-value was used as representative for purposes of visualization (see
1226 Supplementary Note). DEPICT for ExomeChip was written using the Python programming language, and
1227 the code can be found at https://github.com/RebeccaFine/obesity-ec-depict.

1228 **Pathway enrichment analyses: PASCAL**

1229 We also applied the PASCAL pathway analysis tool[23] to exome-wide association summary statistics
1230 from Stage 1 for all coding variants. The method derives gene-based scores (both SUM and MAX statistics)
1231 and subsequently tests for over-representation of high gene scores in predefined biological pathways. We
1232 used standard pathway libraries from KEGG, REACTOME and BIOCARTA, and also added dichotomized (Z-
1233 score>3) reconstituted gene sets from DEPICT[21]. To accurately estimate SNP-by-SNP correlations even for
1234 rare variants, we used the UK10K data (TwinsUK[89] and ALSPAC[90] studies , N=3781). In order to separate
1235 the contribution of regulatory variants from the coding variants, we also applied PASCAL to association
1236 summary statistics of only regulatory variants (20 kb upstream) and regulatory+coding variants from the
1237 Shungin et al[10] study. In this way, we could comment on what is gained by analyzing coding variants
1238 available on ExomeChip arrays. We performed both MAX and SUM estimations for pathway enrichment.
1239 MAX is more sensitive to genesets driven primarily by a single signal, while SUM is better when there are
1240 multiple variant associations in the same gene.

1241 **Monogenic obesity enrichment analyses**

1242 We compiled two lists consisting of 31 genes with strong evidence that disruption causes
1243 monogenic forms of insulin resistance or diabetes; and 8 genes with evidence that disruption causes
1244 monogenic forms of lipodystrophy. To test for enrichment of association, we conducted simulations by
1245 matching each gene with others based on gene length and number of variants tested, to create a matched

1246 set of genes. We generated 1,000 matched gene sets from our data, and assessed how often the number
1247 of variants exceeding set significance thresholds was greater than in our monogenic obesity gene set.

1248 **Variance explained**

1249 We estimated the phenotypic variance explained by the association signals in Stage 1 all
1250 ancestries analyses for men, women, and combined sexes[91]. For each associated region, we pruned
1251 subsets of SNPs within 500 kb, as this threshold was comparable with previous studies, of the SNPs with
1252 the lowest P-value and used varying P value thresholds (ranging from $2x10^{-7}$ to 0.02) from the combined
1253 sexes results. Additionally, we examined all variants and independent variants across a range of MAF
1254 thresholds. The variance explained by each subset of SNPs in each strata was estimated by summing the
1255 variance explained by the individual top coding variants. For the comparison of variance explained
1256 between men and women, we tested for the significance of the differences assuming that the weighted
1257 sum of chi-squared distributed variables tend to a Gaussian distribution ensured by Lyapunov's central
1258 limit theorem.[91,92]

1259 **Cross-trait lookups**

1260 To carefully explore the relationship between WHRadjBMI and related cardiometabolic,
1261 anthropometric, and reproductive traits, association results for the 51 WHRadjBMI coding SNPs were
1262 requested from existing or on-going meta-analyses from 7 consortia, including ExomeChip data from
1263 GIANT (BMI, height), Global Lipids Genetics Consortium Results (GLGC) (total cholesterol, triglycerides,
1264 HDL-cholesterol, LDL-cholesterol), International Consortium for Blood Pressure (IBPC)[93] (systolic and
1265 diastolic blood pressure), Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC)
1266 (glycemic traits), and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (type 2
1267 diabetes).).[22,25-29] For coronary artery disease, we accessed 1000 Genomes Project-imputed GWAS data
1268 released by CARDIoGRAMplusC4D[94] and for the ReproGen consortium (age at menarche and menopause)

1269 we used a combination of ExomeChip and 1000 Genomes Project-Imputed GWAS data. Heatmaps were
1270 generated in R v3.3.2 using gplots (https://CRAN.R-project.org/package=gplots). We used Euclidean
1271 distance based on p-value and direction of effect and complete linkage clustering for the dendrograms.

1272 **GWAS Catalog Lookups**

1273 In order to determine if significant coding variants were associated with any related
1274 cardiometabolic and anthropometric traits, we also searched the NHGRI-EBI GWAS Catalog for previous
1275 variant-trait associations near our lead SNPs (+/- 500 kb). We used PLINK to calculate LD for variants using
1276 ARIC study European participants. All SNVs within the specified regions with an r^2 value > 0.7 were retained
1277 from NHGRI-EBI GWAS Catalog for further evaluation[37]. Consistent direction of effect was based on WHR-
1278 increasing allele, LD, and allele frequency. Therefore, when a GWAS Catalog variant was not identical or
1279 in high LD (r^2 > 0.9) with the WHR variant, and MAF >0.45, we do not comment on direction of effect.

1280 **Body-fat percentage associations**

1281 We performed body fat percent and truncal fat percent look-up of 48 of the 56 identified variants
1282 (tables 1 and 2) that were available in the UK Biobank, Release 1[84], data (notably some of the rare variants
1283 in table 1 and 2 were not available) to further characterize their effects on WHRadjBMI. Genome-wide
1284 association analyses for body fat percent and truncal fat percent were carried out in the UK Biobank. Prior
1285 to analysis, phenotype data were filtered to exclude pregnant or possibly pregnant women, individuals
1286 with body mass index < 15, and without genetically confirmed European ancestry, resulting in a sample
1287 size of 120,286. Estimated measures of body fat percent and truncal fat percent were obtained using the
1288 Tanita BC418MA body composition analyzer (Tanita, Tokyo, Japan). Individuals were not required to fast
1289 and did not follow any specific instructions prior to the bioimpedance measurements. SNPTEST was used
1290 to perform the analyses based on residuals adjusted for age, 15 principle components, assessment center
1291 and the genotyping chip[85].

1292 **Collider bias**

1293 In order to evaluate SNPs for possible collider bias[18], we used results from a recent association
1294 analysis from GIANT on BMI[25]. For each significant SNP identified in our additive models, WHRadjBMI
1295 associations were corrected for potential bias due to associations between each variant and BMI (See
1296 **Supplementary Note 1** for additional details). Variants were considered robust against collider bias if they
1297 met Bonferroni-corrected significance following correction ($P_{corrected}<9.09x10^{-4}$, 0.05/55 variants
1298 examined).

1299 **Drosophila RNAi knockdown experiments**

1300 For each gene in which coding variants were associated with WHRadjBMI in the final combined
1301 meta-analysis ($P < 2\times10^{-7}$), its corresponding Drosophila orthologues were identified in the Ensembl
1302 ortholog database (www.ensembl.org), when available. Drosophila triglyceride content values were
1303 mined from a publicly available genome-wide fat screen data set [45] to identify potential genes for follow-
1304 up knockdowns. Estimated values represent fractional changes in triglyceride content in adult male flies.
1305 Data are from male progeny resulting from crosses of male UAS-RNAi flies from the Vienna Drosophila
1306 Resource Center (VDRC) and Hsp70-GAL4; Tub-GAL8ts virgin females. Two-to-five-day-old males were
1307 sorted into groups of 20 and subjected to two one-hour wet heatshocks four days apart. On the seventh
1308 day, flies were picked in groups of eight, manually crushed and sonicated, and the lysates heat-inactivated
1309 for 10 min in a thermocycler at 95 °C. Centrifuge-cleared supernatants were then used for triglyceride
1310 (GPO Trinder, Sigma) and protein (Pierce) determination. Triglyceride values from these adult-induced
1311 ubiquitous RNAi knockdown individuals were normalized to those obtained in parallel from non-
1312 heatshocked progeny from the very same crosses. The screen comprised one to three biological replicates.
1313 We followed up each gene with a >0.2 increase or >0.4 decrease in triglyceride content.

1314 Orthologues for two genes were brought forward for follow-up, *DNAH10* and *PLXND1*. For both
1315 genes, we generated adipose tissue (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdown crosses

1316 to knockdown transcripts in a tissue specific manner, leveraging upstream activation sequence (UAS)-
1317 inducible short-hairpin knockdown lines, available through the VDRC (Vienna *Drosophila* Resource
1318 Center). Specifically, elav-Gal4, which drives expression of the RNAi construct in post mitotic neurons
1319 starting at embryonic stages all the way to adulthood, was used. Cg drives expression in the fat body and
1320 hemocytes starting at embryonic stage 12, all the way to adulthood. We crossed male UAS-RNAi flies and
1321 elav-GAL4 or CG-GAL4 virgin female flies. All fly experiments were carried out at 25°C. Five-to-seven-day-
1322 old males were sorted into groups of 20, weighed and homogenated in PBS with 0.05% Tween with Lysing
1323 Matrix D in a beadshaker. The homogenate was heat-inactivated for 10 min in a thermocycler at 70°C.
1324 10µl of the homogenate was subsequently used in a triglyceride assay (Sigma, Serum Triglyceride
1325 Determination Kit) which was carried out in duplicate according to protocol, with one alteration: the
1326 samples were cleared of residual particulate debris by centrifugation before absorbance reading.
1327 Resulting triglyceride values were normalized to fly weight and larval/population density. We used the
1328 non-parametric Kruskall-Wallis test to compare wild type with knockdown lines.

1329 **Expression quantitative trait loci (eQTLs) analysis**

1330 We queried the significant variant (Exome coding SNPs)-gene pairs associated with eGenes across
1331 five metabolically relevant tissues (skeletal muscle, subcutaneous adipose, visceral adipose, liver and
1332 pancreas) with at least 70 samples in the GTEx database[46]. For each tissue, variants were selected based
1333 on the following thresholds: the minor allele was observed in at least 10 samples, and the minor allele
1334 frequency was ≥ 0.01. eGenes, genes with a significant eQTL, are defined on a false discovery rate (FDR)[95]
1335 threshold of ≤0.05 of beta distribution-adjusted empirical p-value from FastQTL. Nominal p-values were
1336 generated for each variant-gene pair by testing the alternative hypothesis that the slope of a linear
1337 regression model between genotype and expression deviates from 0. To identify the list of all significant
1338 variant-gene pairs associated with eGenes, a genome-wide empirical p-value threshold[64], pt, was defined
1339 as the empirical p-value of the gene closest to the 0.05 FDR threshold. pt was then used to calculate a

1340 nominal p-value threshold for each gene based on the beta distribution model (from FastQTL) of the
1341 minimum p-value distribution f(pmin) obtained from the permutations for the gene. For each gene,
1342 variants with a nominal p-value below the gene-level threshold were considered significant and included
1343 in the final list of variant-gene pairs[64]. For each eGene, we also listed the most significantly associated
1344 variants (eSNP). Only these exome SNPs with r^2 > 0.8 with eSNPs were considered for the biological
1345 interpretation (Supplementary eQTL GTEx).

1346 We also performed cis-eQTL analysis in 770 METSIM subcutaneous adipose tissue samples as
1347 described in Civelek, et al.[96] A false discovery rate (FDR) was calculated using all p-values from the cis-
1348 eQTL detection in the q-value package in R. Variants associated with nearby genes at an FDR less than 1%
1349 were considered to be significant (equivalent p-value < 2.46×10^{-4}).

1350 For loci with more than one microarray probeset of the same gene associated with the exome
1351 variant, we selected the probeset that provided the strongest LD r2 between the exome variant and the
1352 eSNP. In reciprocal conditional analysis, we conditioned on the lead exome variant by including it as a
1353 covariate in the cis-eQTL detection and reporting the p-value of the eSNP and vice versa. We considered
1354 the signals to be coincident if both the lead exome variant and the eSNP were no longer significant after
1355 conditioning on the other and the variants were in high pairwise LD (r2 > 0.80).

1356 For loci that also harbored reported GWAS variants, we performed reciprocal conditional analysis
1357 between the GWAS lead variant and the lead eSNP. For loci with more than one reported GWAS variant,
1358 the GWAS lead variant with the strongest LD r2 with the lead eSNP was reported.

1359 **Penetrance analysis**

1360 Phenotype and genotype data from the UK Biobank (UKBB) were used for the penetrance analysis.
1361 Three of 16 rare and low frequency variants (MAF ≤ 1%) detected in the final Stage 1 plus 2 meta-analysis
1362 were available in the UKBB and had relatively larger effect sizes (>0.90). The phenotype data for these

1363 three variants were stratified with respect to waist-to-hip ratio (WHR) using the World Health

1364 Organization (WHO) guidelines. These guidelines consider women and men with WHR greater than 0.85

1365 and 0.90 as obese, respectively. Genotype and allele counts were obtained for the available variants and

1366 these were used to calculate the number of carriers of the minor allele. The number of carriers for women,

1367 men and all combined was then compared between two strata (obese vs. non-obese) using a $\chi 2$ test. The

1368 significance threshold was determined by using a Bonferroni correction for the number of tests performed

1369 ($0.05/9=5.5 \times 10^{-3}$)).

1370 DATA AVAILABILITY

1371 Summary statistics of all analyses are available at https://www.broadinstitute.org/collaboration/giant/.

1372

1373 **BOXES**

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=9.16×10^{-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 system[97-101]. Mutations in this gene are associated with Moebius syndrome[102-105], and persistent truncus arteriosus[99,106]. *PLXND1* is involved in angiogenesis as part of the SEMA and VEGF signalling pathways[107-110]. *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 mellitus[101]. *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= 4.81×10^{-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 differentiation[64-68], including adipocytes[68]. 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 individuals[70]. *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 percentile[112]. *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 sensitivity[115]. Angptl4-deficient mice have

hypotriglyceridemia and increased lipoprotein lipase (LPL) activity, while transgenic mice overexpressing Angplt4 in the liver have higher plasma triglyceride levels and decreased LPL activity[116]. The major allele of rs116843064 has been previously associated with increased risk of coronary heart disease and increased TG[63]. *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=3.9×10^{-15}). This variant is independent of the previously reported GWAS signal in the *RREB1* region (rs1294410; 6:6738752[10]). 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 differentiation[117-119]. The ras responsive transcription factor *RREB1* is a candidate gene for type 2 diabetes associated end-stage kidney disease[118]. 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=6.24×10^{-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, r^2= 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 pathway [121]. *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 secretion[124], may be important in regulating adipocyte differentiation[125-127], plays an important role in regulating adiposity and energy balance[128]. *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 individuals[129]. *TBX15* negatively controls depot-specific adipocyte differentiation and function[130] and regulates glycolytic myofiber identity and muscle metabolism[131]. *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).

1374 **REFERENCES**

1375 1. Pischon, T. *et al.* General and abdominal adiposity and risk of death in Europe. *N Engl J Med* **359**,
1376 2105-20 (2008).
1377 2. Wang, Y., Rimm, E.B., Stampfer, M.J., Willett, W.C. & Hu, F.B. Comparison of abdominal adiposity
1378 and overall obesity in predicting risk of type 2 diabetes among men. *Am J Clin Nutr* **81**, 555-63
1379 (2005).
1380 3. Canoy, D. Distribution of body fat and risk of coronary heart disease in men and women. *Curr Opin*
1381 *Cardiol* **23**, 591-8 (2008).
1382 4. Snijder, M.B. *et al.* Associations of hip and thigh circumferences independent of waist
1383 circumference with the incidence of type 2 diabetes: the Hoorn Study. *Am J Clin Nutr* **77**, 1192-7
1384 (2003).
1385 5. Yusuf, S. *et al.* Obesity and the risk of myocardial infarction in 27,000 participants from 52
1386 countries: a case-control study. *Lancet* **366**, 1640-9 (2005).
1387 6. Mason, C., Craig, C.L. & Katzmarzyk, P.T. Influence of central and extremity circumferences on all-
1388 cause mortality in men and women. *Obesity (Silver Spring)* **16**, 2690-5 (2008).
1389 7. Karpe, F. & Pinnick, K.E. Biology of upper-body and lower-body adipose tissue--link to whole-body
1390 phenotypes. *Nat Rev Endocrinol* **11**, 90-100 (2015).
1391 8. Manolopoulos, K.N., Karpe, F. & Frayn, K.N. Gluteofemoral body fat as a determinant of metabolic
1392 health. *Int J Obes (Lond)* **34**, 949-59 (2010).
1393 9. Emdin, C.A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2
1394 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626-634 (2017).
1395 10. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature*
1396 **518**, 187-96 (2015).

1397 11. Winkler, T.W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size
1398 and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
1399 12. Wen, W. *et al.* Genome-wide association studies in East Asians identify new loci for waist-hip ratio
1400 and waist circumference. *Sci Rep* **6**, 17958 (2016).
1401 13. Gao, C. *et al.* A Comprehensive Analysis of Common and Rare Variants to Identify Adiposity Loci
1402 in Hispanic Americans: The IRAS Family Study (IRASFS). *PLoS One* **10**, e0134649 (2015).
1403 14. Graff, M. *et al.* Genome-wide physical activity interactions in adiposity - A meta-analysis of
1404 200,452 adults. *PLoS Genet* **13**, e1006528 (2017).
1405 15. Justice, A.E. *et al.* Genome-wide meta-analysis of 241,258 adults accounting for smoking
1406 behaviour identifies novel loci for obesity traits. *Nat Commun* **8**, 14977 (2017).
1407 16. Ng, M.C.Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of
1408 genome-wide association studies in individuals of African ancestry: African Ancestry
1409 Anthropometry Genetics Consortium. *PLoS Genet* **13**, e1006719 (2017).
1410 17. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature*
1411 **518**, 197-206 (2015).
1412 18. Aschard, H., Vilhjalmsson, B.J., Joshi, A.D., Price, A.L. & Kraft, P. Adjusting for heritable covariates
1413 can bias effect estimates in genome-wide association studies. *Am J Hum Genet* **96**, 329-39 (2015).
1414 19. Day, F.R., Loh, P.R., Scott, R.A., Ong, K.K. & Perry, J.R. A Robust Example of Collider Bias in a
1415 Genetic Association Study. *Am J Hum Genet* **98**, 392-3 (2016).
1416 20. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-
1417 analysis for rare variants. *Bioinformatics* **30**, 2828-9 (2014).
1418 21. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene
1419 functions. *Nat Commun* **6**, 5890 (2015).

1420 22. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**,
1421 186-190 (2017).
1422 23. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation
1423 of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput Biol* **12**, e1004714
1424 (2016).
1425 24. Kawai, M., de Paula, F.J. & Rosen, C.J. New insights into osteoporosis: the bone-fat connection. *J*
1426 *Intern Med* **272**, 317-29 (2012).
1427 25. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that
1428 control energy intake and expenditure in obesity. *Nat Genet* **50**, 26-41 (2018).
1429 26. Liu, D.J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. **49**, 1758-
1430 1766 (2017).
1431 27. Kraja, A.T. *et al.* New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000
1432 Individuals. *Circ Cardiovasc Genet* **10**(2017).
1433 28. Mahajan, A. *et al.* Identification and functional characterization of G6PC2 coding variants
1434 influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet*
1435 **11**, e1004876 (2015).
1436 29. Manning, A. *et al.* A Low-Frequency Inactivating AKT2 Variant Enriched in the Finnish Population
1437 Is Associated With Fasting Insulin Levels and Type 2 Diabetes Risk. *Diabetes* **66**, 2019-2032 (2017).
1438 30. Zhao, W. *et al.* Identification of new susceptibility loci for type 2 diabetes and shared etiological
1439 pathways with coronary heart disease. **49**, 1450-1457 (2017).
1440 31. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture
1441 and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
1442 32. Ng, M.C. *et al.* Meta-analysis of genome-wide association studies in African Americans provides
1443 insights into the genetic architecture of type 2 diabetes. *PLoS Genet* **10**, e1004517 (2014).

1444 33. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic
1445 architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-44 (2014).
1446 34. Saxena, R. *et al.* Genome-wide association study identifies a novel locus contributing to type 2
1447 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* **62**, 1746-55 (2013).
1448 35. Cook, J.P. & Morris, A.P. Multi-ethnic genome-wide association study identifies novel locus for
1449 type 2 diabetes susceptibility. *Eur J Hum Genet* **24**, 1175-80 (2016).
1450 36. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale
1451 association analysis. *Nat Genet* **42**, 579-89 (2010).
1452 37. Burdett, T. *et al.* The NHGRI-EBI Catalog of published genome-wide association studies. v1.0 edn
1453 Vol. 2015 (2015).
1454 38. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci
1455 for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
1456 39. Lutoslawska, G. *et al.* Relationship between the percentage of body fat and surrogate indices of
1457 fatness in male and female Polish active and sedentary students. *J Physiol Anthropol* **33**, 10 (2014).
1458 40. Verma, M., Rajput, M., Sahoo, S.S., Kaur, N. & Rohilla, R. Correlation between the percentage of
1459 body fat and surrogate indices of obesity among adult population in rural block of Haryana. *J
1460 Family Med Prim Care* **5**, 154-9 (2016).
1461 41. Pereira, P.F. *et al.* [Measurements of location of body fat distribution: an assessment of colinearity
1462 with body mass, adiposity and stature in female adolescents]. *Rev Paul Pediatr* **33**, 63-71 (2015).
1463 42. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic
1464 disease risk. *Nat Commun* **7**, 10495 (2016).
1465 43. Chambers, J.C. *et al.* Common genetic variation near MC4R is associated with waist circumference
1466 and insulin resistance. *Nat Genet* **40**, 716-8 (2008).

1467 44. Nead, K.T. *et al.* Contribution of common non-synonymous variants in PCSK1 to body mass index
1468 variation and risk of obesity: a systematic review and meta-analysis with evidence from up to 331
1469 175 individuals. *Hum Mol Genet* **24**, 3582-94 (2015).
1470 45. Pospisilik, J.A. *et al.* Drosophila genome-wide obesity screen reveals hedgehog as a determinant
1471 of brown versus white adipose cell fate. *Cell* **140**, 148-60 (2010).
1472 46. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
1473 multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
1474 47. Baraille, F., Planchais, J., Dentin, R., Guilmeau, S. & Postic, C. Integration of ChREBP-Mediated
1475 Glucose Sensing into Whole Body Metabolism. *Physiology (Bethesda)* **30**, 428-37 (2015).
1476 48. Kursawe, R. *et al.* Decreased transcription of ChREBP-alpha/beta isoforms in abdominal
1477 subcutaneous adipose tissue of obese adolescents with prediabetes or early type 2 diabetes:
1478 associations with insulin resistance and hyperglycemia. *Diabetes* **62**, 837-44 (2013).
1479 49. Lotta, L.A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage
1480 capacity in the pathogenesis of human insulin resistance. *Nat Genet* **49**, 17-26 (2017).
1481 50. Cargill, M. *et al.* A large-scale genetic association study confirms IL12B and leads to the
1482 identification of IL23R as psoriasis-risk genes. *Am J Hum Genet* **80**, 273-90 (2007).
1483 51. Hazlett, J., Stamp, L.K., Merriman, T., Highton, J. & Hessian, P.A. IL-23R rs11209026 polymorphism
1484 modulates IL-17A expression in patients with rheumatoid arthritis. *Genes Immun* **13**, 282-7 (2012).
1485 52. Karaderi, T. *et al.* Association between the interleukin 23 receptor and ankylosing spondylitis is
1486 confirmed by a new UK case-control study and meta-analysis of published series. *Rheumatology*
1487 *(Oxford)* **48**, 386-9 (2009).
1488 53. Duerr, R.H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel
1489 disease gene. *Science* **314**, 1461-3 (2006).

1490 54. Abdollahi, E., Tavasolian, F., Momtazi-Borojeni, A.A., Samadi, M. & Rafatpanah, H. Protective role
1491 of R381Q (rs11209026) polymorphism in IL-23R gene in immune-mediated diseases: A
1492 comprehensive review. *J Immunotoxicol* **13**, 286-300 (2016).
1493 55. Abraham, C., Dulai, P.S., Vermeire, S. & Sandborn, W.J. Lessons Learned From Trials Targeting
1494 Cytokine Pathways in Patients With Inflammatory Bowel Diseases. *Gastroenterology* **152**, 374-388
1495 e4 (2017).
1496 56. Molinelli, E., Campanati, A., Ganzetti, G. & Offidani, A. Biologic Therapy in Immune Mediated
1497 Inflammatory Disease: Basic Science and Clinical Concepts. *Curr Drug Saf* **11**, 35-43 (2016).
1498 57. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-7 (2016).
1499 58. Wells, J.C. Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* **21**, 415-
1500 30 (2007).
1501 59. Loomba-Albrecht, L.A. & Styne, D.M. Effect of puberty on body composition. *Curr Opin Endocrinol*
1502 *Diabetes Obes* **16**, 10-5 (2009).
1503 60. Rogol, A.D., Roemmich, J.N. & Clark, P.A. Growth at puberty. *J Adolesc Health* **31**, 192-200 (2002).
1504 61. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-45 (2012).
1505 62. Stern, J.H., Rutkowski, J.M. & Scherer, P.E. Adiponectin, Leptin, and Fatty Acids in the
1506 Maintenance of Metabolic Homeostasis through Adipose Tissue Crosstalk. *Cell Metab* **23**, 770-84
1507 (2016).
1508 63. Dewey, F.E. *et al.* Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J*
1509 *Med* **374**, 1123-33 (2016).
1510 64. Bondestam, J. *et al.* cDNA cloning, expression studies and chromosome mapping of human type I
1511 serine/threonine kinase receptor ALK7 (ACVR1C). *Cytogenet Cell Genet* **95**, 157-62 (2001).

1512 65. Jornvall, H., Blokzijl, A., ten Dijke, P. & Ibanez, C.F. The orphan receptor serine/threonine kinase
1513 ALK7 signals arrest of proliferation and morphological differentiation in a neuronal cell line. *J Biol*
1514 *Chem* **276**, 5140-6 (2001).
1515 66. Kim, B.C. *et al.* Activin receptor-like kinase-7 induces apoptosis through activation of MAPKs in a
1516 Smad3-dependent mechanism in hepatoma cells. *J Biol Chem* **279**, 28458-65 (2004).
1517 67. Watanabe, R. *et al.* The MH1 domains of smad2 and smad3 are involved in the regulation of the
1518 ALK7 signals. *Biochem Biophys Res Commun* **254**, 707-12 (1999).
1519 68. Kogame, M. *et al.* ALK7 is a novel marker for adipocyte differentiation. *J Med Invest* **53**, 238-45
1520 (2006).
1521 69. Murakami, M. *et al.* Expression of activin receptor-like kinase 7 in adipose tissues. *Biochem Genet*
1522 **51**, 202-10 (2013).
1523 70. Carlsson, L.M. *et al.* ALK7 expression is specific for adipose tissue, reduced in obesity and
1524 correlates to factors implicated in metabolic disease. *Biochem Biophys Res Commun* **382**, 309-14
1525 (2009).
1526 71. Carithers, L.J. & Moore, H.M. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank*
1527 **13**, 307-8 (2015).
1528 72. Yogosawa, S., Mizutani, S., Ogawa, Y. & Izumi, T. Activin receptor-like kinase 7 suppresses lipolysis
1529 to accumulate fat in obesity through downregulation of peroxisome proliferator-activated
1530 receptor gamma and C/EBPalpha. *Diabetes* **62**, 115-23 (2013).
1531 73. Yogosawa, S. & Izumi, T. Roles of activin receptor-like kinase 7 signaling and its target, peroxisome
1532 proliferator-activated receptor gamma, in lean and obese adipocytes. *Adipocyte* **2**, 246-50 (2013).
1533 74. Seifi, M., Ghasemi, A., Namipashaki, A. & Samadikuchaksaraei, A. Is C771G polymorphism of MLX
1534 interacting protein-like (MLXIPL) gene a novel genetic risk factor for non-alcoholic fatty liver
1535 disease? *Cell Mol Biol (Noisy-le-grand)* **60**, 37-42 (2014).

1536 75. Cairo, S., Merla, G., Urbinati, F., Ballabio, A. & Reymond, A. WBSCR14, a gene mapping to the
1537 Williams--Beuren syndrome deleted region, is a new member of the Mlx transcription factor
1538 network. *Hum Mol Genet* **10**, 617-27 (2001).
1539 76. Ambele, M.A., Dessels, C., Durandt, C. & Pepper, M.S. Genome-wide analysis of gene expression
1540 during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene
1541 expression during adipocyte differentiation. *Stem Cell Res* **16**, 725-34 (2016).
1542 77. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet* **46**, 200-4
1543 (2014).
1544 78. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population
1545 analysis. *Bioinformatics* **28**, 2543-5 (2012).
1546 79. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat*
1547 *Protoc* **9**, 1192-212 (2014).
1548 80. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature*
1549 **518**, 187-196 (2015).
1550 81. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**,
1551 185-90 (2014).
1552 82. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12
1553 (2011).
1554 83. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies
1555 additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
1556 84. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range
1557 of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
1558 85. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for
1559 genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

1560 86. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven
1561 common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
1562 87. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev*
1563 *Genet* **11**, 499-511 (2010).
1564 88. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6
1565 (2007).
1566 89. Moayyeri, A., Hammond, C.J., Valdes, A.M. & Spector, T.D. Cohort Profile: TwinsUK and healthy
1567 ageing twin study. *Int J Epidemiol* **42**, 76-85 (2013).
1568 90. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal
1569 Study of Parents and Children. *Int J Epidemiol* **42**, 111-27 (2013).
1570 91. Kutalik, Z., Whittaker, J., Waterworth, D., Beckmann, J.S. & Bergmann, S. Novel method to
1571 estimate the phenotypic variation explained by genome-wide association studies reveals large
1572 fraction of the missing heritability. *Genet Epidemiol* **35**, 341-9 (2011).
1573 92. Billingsley, P. *Probability and measure*, xii, 622 p. (Wiley, New York, 1986).
1574 93. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated
1575 with blood pressure and hypertension. *Nat Genet* **48**, 1151-61 (2016).
1576 94. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis
1577 of coronary artery disease. *Nat Genet* **47**, 1121-30 (2015).
1578 95. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S*
1579 *A* **100**, 9440-5 (2003).
1580 96. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am*
1581 *J Hum Genet* **100**, 428-443 (2017).
1582 97. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**, D222-6
1583 (2015).

1584 98. Toyofuku, T. *et al.* Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses
1585 angiogenesis via Plexin-D1. *EMBO J* **26**, 1373-84 (2007).
1586 99. Gitler, A.D., Lu, M.M. & Epstein, J.A. PlexinD1 and semaphorin signaling are required in endothelial
1587 cells for cardiovascular development. *Dev Cell* **7**, 107-16 (2004).
1588 100. Luchino, J. *et al.* Semaphorin 3E suppresses tumor cell death triggered by the plexin D1
1589 dependence receptor in metastatic breast cancers. *Cancer Cell* **24**, 673-85 (2013).
1590 101. Shimizu, I. *et al.* Semaphorin3E-induced inflammation contributes to insulin resistance in dietary
1591 obesity. *Cell Metab* **18**, 491-504 (2013).
1592 102. Verzijl, H.T., van der Zwaag, B., Cruysberg, J.R. & Padberg, G.W. Mobius syndrome redefined: a
1593 syndrome of rhombencephalic maldevelopment. *Neurology* **61**, 327-33 (2003).
1594 103. Verzijl, H.T., van der Zwaag, B., Lammens, M., ten Donkelaar, H.J. & Padberg, G.W. The
1595 neuropathology of hereditary congenital facial palsy vs Mobius syndrome. *Neurology* **64**, 649-53
1596 (2005).
1597 104. Fujita, M., Reinhart, F. & Neutra, M. Convergence of apical and basolateral endocytic pathways at
1598 apical late endosomes in absorptive cells of suckling rat ileum in vivo. *J Cell Sci* **97 (Pt 2)**, 385-94
1599 (1990).
1600 105. Briegel, W. Neuropsychiatric findings of Mobius sequence -- a review. *Clin Genet* **70**, 91-7 (2006).
1601 106. Ta-Shma, A. *et al.* Isolated truncus arteriosus associated with a mutation in the plexin-D1 gene.
1602 *Am J Med Genet A* **161A**, 3115-20 (2013).
1603 107. Mazzotta, C. *et al.* Plexin-D1/Semaphorin 3E pathway may contribute to dysregulation of vascular
1604 tone control and defective angiogenesis in systemic sclerosis. *Arthritis Res Ther* **17**, 221 (2015).
1605 108. Yang, W.J. *et al.* Semaphorin-3C signals through Neuropilin-1 and PlexinD1 receptors to inhibit
1606 pathological angiogenesis. *EMBO Mol Med* **7**, 1267-84 (2015).

1607 109. Zygmunt, T. *et al.* Semaphorin-PlexinD1 signaling limits angiogenic potential via the VEGF decoy
1608 receptor sFlt1. *Dev Cell* **21**, 301-14 (2011).
1609 110. Kim, J., Oh, W.J., Gaiano, N., Yoshida, Y. & Gu, C. Semaphorin 3E-Plexin-D1 signaling regulates
1610 VEGF function in developmental angiogenesis via a feedback mechanism. *Genes Dev* **25**, 1399-411
1611 (2011).
1612 111. Bertolino, P. *et al.* Activin B receptor ALK7 is a negative regulator of pancreatic beta-cell function.
1613 *Proc Natl Acad Sci U S A* **105**, 7246-51 (2008).
1614 112. Haworth, K.E. *et al.* Methylation of the FGFR2 gene is associated with high birth weight centile in
1615 humans. *Epigenomics* **6**, 477-91 (2014).
1616 113. Chi, X. *et al.* Angiopoietin-like 4 Modifies the Interactions between Lipoprotein Lipase and Its
1617 Endothelial Cell Transporter GPIHBP1. *J Biol Chem* **290**, 11865-77 (2015).
1618 114. Catoire, M. *et al.* Fatty acid-inducible ANGPTL4 governs lipid metabolic response to exercise. *Proc*
1619 *Natl Acad Sci U S A* **111**, E1043-52 (2014).
1620 115. van Raalte, D.H. *et al.* Angiopoietin-like protein 4 is differentially regulated by glucocorticoids and
1621 insulin in vitro and in vivo in healthy humans. *Exp Clin Endocrinol Diabetes* **120**, 598-603 (2012).
1622 116. Koster, A. *et al.* Transgenic angiopoietin-like (angptl)4 overexpression and targeted disruption of
1623 angptl4 and angptl3: regulation of triglyceride metabolism. *Endocrinology* **146**, 4943-50 (2005).
1624 117. Thiagalingam, A. *et al.* RREB-1, a novel zinc finger protein, is involved in the differentiation
1625 response to Ras in human medullary thyroid carcinomas. *Mol Cell Biol* **16**, 5335-45 (1996).
1626 118. Bonomo, J.A. *et al.* The ras responsive transcription factor RREB1 is a novel candidate gene for
1627 type 2 diabetes associated end-stage kidney disease. *Hum Mol Genet* **23**, 6441-7 (2014).
1628 119. Thiagalingam, A., Lengauer, C., Baylin, S.B. & Nelkin, B.D. RREB1, a ras responsive element binding
1629 protein, maps to human chromosome 6p25. *Genomics* **45**, 630-2 (1997).

1630 120. Bisogno, T. *et al.* Cloning of the first sn1-DAG lipases points to the spatial and temporal regulation
1631 of endocannabinoid signaling in the brain. *J Cell Biol* **163**, 463-8 (2003).
1632 121. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat*
1633 *Genet* **45**, 1274-83 (2013).
1634 122. Kooner, J.S. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma
1635 triglycerides. *Nat Genet* **40**, 149-51 (2008).
1636 123. Pan, L.A. *et al.* G771C Polymorphism in the MLXIPL Gene Is Associated with a Risk of Coronary
1637 Artery Disease in the Chinese: A Case-Control Study. *Cardiology* **114**, 174-8 (2009).
1638 124. Kang, G., Leech, C.A., Chepurny, O.G., Coetzee, W.A. & Holz, G.G. Role of the cAMP sensor Epac
1639 as a determinant of KATP channel ATP sensitivity in human pancreatic beta-cells and rat INS-1
1640 cells. *J Physiol* **586**, 1307-19 (2008).
1641 125. Ji, Z., Mei, F.C. & Cheng, X. Epac, not PKA catalytic subunit, is required for 3T3-L1 preadipocyte
1642 differentiation. *Front Biosci (Elite Ed)* **2**, 392-8 (2010).
1643 126. Martini, C.N., Plaza, M.V. & Vila Mdel, C. PKA-dependent and independent cAMP signaling in 3T3-
1644 L1 fibroblasts differentiation. *Mol Cell Endocrinol* **298**, 42-7 (2009).
1645 127. Petersen, R.K. *et al.* Cyclic AMP (cAMP)-mediated stimulation of adipocyte differentiation requires
1646 the synergistic action of Epac- and cAMP-dependent protein kinase-dependent processes. *Mol*
1647 *Cell Biol* **28**, 3804-16 (2008).
1648 128. Yan, J. *et al.* Enhanced leptin sensitivity, reduced adiposity, and improved glucose homeostasis in
1649 mice lacking exchange protein directly activated by cyclic AMP isoform 1. *Mol Cell Biol* **33**, 918-26
1650 (2013).
1651 129. Gesta, S. *et al.* Evidence for a role of developmental genes in the origin of obesity and body fat
1652 distribution. *Proc Natl Acad Sci U S A* **103**, 6676-81 (2006).

1653 130. Gesta, S. *et al.* Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and
1654 mitochondrial respiration. *Proc Natl Acad Sci U S A* **108**, 2771-6 (2011).
1655 131. Lee, K.Y. *et al.* Tbx15 controls skeletal muscle fibre-type determination and muscle metabolism.
1656 *Nat Commun* **6**, 8054 (2015).
1657
1658
1659

1660 **FIGURES**

1661 **Figure 1. Summary of meta-analysis study design and workflow.** Abbreviations:

1662 EUR- European, AFR- African, SAS- South Asian, EAS- East Asian, and HIS- Hispanic/Latino ancestry.

1663 **Figure 2**. Minor allele frequency compared to estimated effect. This scatter plot displays the relationship

1664 between minor allele frequency (MAF) and the estimated effect (β) for each significant coding variant in

1665 our meta-analyses. All novel WHRadjBMI variants are highlighted in orange, and variants identified only

1666 in models that assume recessive inheritance are denoted by diamonds and only in sex-specific analyses

1667 by triangles. Eighty percent power was calculated based on the total sample size in the Stage 1+2 meta-

1668 analysis and P=2x10^{-7}. Estimated effects are shown in original units (cm/cm) calculated by using effect

1669 sizes in standard deviation (SD) units times SD of WHR in the ARIC study (sexes combined=0.067,

1670 men=0.052, women=0.080).

1671 **Figure 3.** Regional association plots for known loci with novel coding signals. Point color reflects r^2

1672 calculated from the ARIC dataset. In a) there are two independent variants in *RSPO3* and *KIAA0408,* as

1673 shown by conditional analysis. In b) we have a variant in *RREB1* that is independent of the GWAS variant

1674 rs1294421.

1675 **Figure 4.** Heat maps showing DEPICT gene set enrichment results. For any given square, the color indicates

1676 how strongly the corresponding gene (shown on the x-axis) is predicted to belong to the reconstituted

1677 gene set (y-axis). This value is based on the gene's z-score for gene set inclusion in DEPICT's reconstituted

1678 gene sets, where red indicates a higher and blue a lower z-score. To visually reduce redundancy and

1679 increase clarity, we chose one representative "meta-gene set" for each group of highly correlated gene

1680 sets based on affinity propagation clustering (**Online Methods**, **Supplementary Note 2**). Heatmap

1681 intensity and DEPICT P-values (see P-values in **Supplementary Data 4-5**) correspond to the most

1682 significantly enriched gene set within the meta-gene set. Annotations for the genes indicate (1) the minor

1683 allele frequency of the significant ExomeChip (EC) variant (shades of blue; if multiple variants, the lowest-
1684 frequency variant was kept), (2) whether the variant's P-value reached array-wide significance (<2x10-7)
1685 or suggestive significance (<5x10-4) (shades of purple), (3) whether the variant was novel, overlapping
1686 "relaxed" GWAS signals from Shungin et al.[10] (GWAS P<5x10-4), or overlapping "stringent" GWAS signals
1687 (GWAS P<5x10-8) (shades of pink), and (4) whether the gene was included in the gene set enrichment
1688 analysis or excluded by filters (shades of brown/orange) (Online Methods and Supplementary
1689 Information). Annotations for the gene sets indicate if the meta-gene set was found significant (shades of
1690 green; FDR <0.01, <0.05, or not significant) in the DEPICT analysis of GWAS results from Shungin et al.

1691

1692

1693 **TABLES**

1694 **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<2x10-07) in the sex-combined

1695 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	*ACVR1C*	N150H	-	455,526	0.989	0.065	0.011	***4.8E-10***	**1.7E-07**	
2	3:50597092	rs1034405	G	A	*C3orf18*	A162V	-	455,424	0.135	0.016	0.003	***1.9E-07***	8.8E-01	G,C
3	4:120528327	rs3733526	G	A	*PDE5A*	A41V	-	461,521	0.187	0.015	0.003	***2.6E-08***	5.2E-03	
4	6:26108117	rs146860658	T	C	*HIST1H1T*	A69T	-	217,995	0.001	0.229	0.042	***4.3E-08***	6.3E-01	S
5	7:6449496	rs2303361	C	T	*DAGLB*	Q664R	-	475,748	0.221	0.014	0.003	***6.2E-08***	3.4E-03	G
6	10:123279643	rs138315382	T	C	*FGFR2*	synonymous	-	236,962	0.001	0.258	0.049	***1.4E-07***	1.1E-01	G,S
7	11:65403651	rs7114037	C	A	*PCNXL3*	H1822Q	-	448,861	0.954	0.029	0.005	***1.8E-08***	4.4E-01	
8	12:48143315	rs145878042	A	G	*RAPGEF3*	L300P	-	470,513	0.990	0.085	0.010	***7.2E-17***	7.3E-03	
9	12:108618630	rs3764002	C	T	*WSCD2*	T266I	-	474,637	0.737	0.014	0.002	***9.8E-10***	5.5E-01	
10	15:42032383	rs17677991	G	C	*MGA*	P1523A	-	469,874	0.345	0.015	0.002	***3.5E-11***	9.1E-01	
11	16:4432029	rs3810818	A	C	*VASN*	E384A	-	424,163	0.231	0.016	0.003	***2.0E-09***	3.3E-01	
	16:4445327	rs3747579	C	T	*CORO7*	R193Q	-	453,078	0.299	0.018	0.002	***2.2E-13***	4.3E-02	
	16:4484396	rs1139653	A	T	*DNAJA3*	N75Y	-	434,331	0.284	0.015	0.002	***4.3E-10***	1.4E-01	
12	19:49232226	rs2287922	A	G	*RASIP1*	R601C	-	430,272	0.494	0.014	0.002	***1.6E-09***	3.7E-02	
	19:49244220	rs2307019	G	A	*IZUMO1*	A333V	-	476,147	0.558	0.012	0.002	***4.7E-08***	3.9E-02	
13	20:42965811	rs144098855	T	C	*R3HDML*	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	*DARS2*	K196R	-	352,646	0.001	0.201	0.038	***1.4E-07***	6.0E-02	G
15	14:58838668	rs1051860	A	G	*ARID4A*	synonymous	-	367,079	0.411	0.013	0.002	***2.2E-08***	1.3E-01	
16	15:42115747	rs3959569	C	G	*MAPKBP1*	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	*TBX15*	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	*ZBTB7B*	P190S	rs905938	476,440	0.976	0.037	0.007	***3.8E-08***	**7.9E-07**	C
3	2:165551201	rs7607980	T	C	*COBLL1*	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	*TFPI*	N221S	rs1569135	452,638	0.697	0.016	0.002	***3.0E-12***	6.3E-01	
5	3:52558008	rs13303	T	C	*STAB1*	M113T	rs2276824	470,111	0.445	0.019	0.002	***5.5E-18***	6.7E-02	
	3:52833805	rs3617	C	A	*ITIH3*	Q315K		452,150	0.541	0.015	0.002	***1.6E-12***	4.0E-01	C
6	3:129137188	rs62266958	C	T	*EFCAB12*	R197H	rs10804591	476,382	0.936	0.036	0.004	***8.3E-17***	**9.3E-05**	
	3:129284818	rs2625973	A	C	*PLXND1*	L1412V		476,338	0.733	0.016	0.002	***9.2E-11***	**1.6E-05**	
7	4:89625427	rs1804080	G	C	*HERC3*	E946Q	rs9991328	446,080	0.838	0.021	0.003	***1.5E-12***	**4.1E-06**	
	4:89668859	rs7657817	C	T	*FAM13A*	V443I		476,383	0.815	0.016	0.003	***5.0E-09***	**9.6E-05**	
8	5:176516631	rs1966265	A	G	*FGFR4*	V10I	rs6556301	455,246	0.236	0.023	0.003	***1.7E-19***	2.1E-01	
9	6:7211818	**rs1334576**[g]	G	A	*RREB1*	G195R	rs1294410	451,044	0.565	0.017	0.002	***3.9E-15***	1.5E-01	
10	6:34827085	rs9469913	A	T	*UHRF1BP1*	Q984H	rs1776897	309,684	0.847	0.021	0.004	***1.2E-08***	2.7E-01	C
11	6:127476516	rs1892172	A	G	*RSPO3*	synonymous	rs11961815, rs72959041, rs1936805	476,358	0.543	0.031	0.002	***2.6E-47***	**7.7E-09**	
	6:127767954	**rs139745911**[g]	A	G	*KIAA0408*	P504S		391,469	0.010	0.103	0.012	***6.8E-19***	**2.0E-04**	
12	7:73012042	rs35332062	G	A	*MLXIPL*	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	*PLCE1*	R240L	rs10786152	476,475	0.173	0.018	0.003	***2.5E-11***	5.9E-01	
14	11:64031241	rs35169799	T	C	*PLCB3*	S778L	rs11231693	476,457	0.061	0.034	0.004	***9.1E-15***	**1.3E-04**	
15	12:123444507	rs58843120	G	T	*ABDB9*	F92L	rs4765219, rs863750	466,498	0.987	0.053	0.009	***1.3E-08***	3.5E-01	
	12:124265687	rs11057353	T	C	*DNAH10*	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	*CCDC92*	S53C		467,649	0.695	0.029	0.002	***7.3E-37***	**5.5E-11**	
16	15:56756285	rs1715919	G	T	*MNS1*	Q55P	rs8030605	476,274	0.096	0.023	0.004	***8.8E-11***	2.7E-02	
17	16:67397580	rs9922085	G	C	*LRRC36*	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	*IFI30*	R76Q	rs12608504	476,389	0.257	0.015	0.002	***3.5E-10***	3.1E-03	
	19:18304700	rs874628	G	A	*MPV17L2*	M72V		476,388	0.271	0.015	0.002	***1.2E-10***	2.5E-03	
19	20:33971914	rs4911494	T	C	*UQCC1*	R51Q	rs224333	451,064	0.602	0.018	0.002	***2.5E-16***	1.5E-03	
	20:34022387	rs224331	A	C	*GDF5*	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	*PEMT*	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	*PLXND1*	M870V	rs10804591	420,520	0.620	0.014	0.002	***3.1E-09***	**1.6E-04**	

1696 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;
1697 EAF=effect allele frequency; EA=effect allele; OA=other allele.
1698 a Coding variants refer to variants located in the exons and splicing junction regions.
1699 b Variant positions are reported according to Human assembly build 37 and their alleles are coded based on the positive strand.
1700 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).
1701 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].
1702 e Effect size is based on standard deviation (SD) per effect allele
1703 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
1704 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).
1705 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,
1706 rs1936805, in a known locus (see Supplementary 8A/B).

1707 h 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 available in stage 2 studies for validation
1708 of Stage 1 association.
1709

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<2x10-07) in the sex-specific meta-analyses and reach bonferonni corrected P-value for sex hetergeneity (P_{sexhet}<7.14E-04).

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	*UGGT2*	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	*MMP14*	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	*DSTYK*	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	*ACVR1C*	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	*ANGPTL4*	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	*ZBTB7B*	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	*COBLL1*	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	*EFCAB12*	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	*PLXND1*	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	*HERC3*	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	*FAM13A*	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	*RSPO3*	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[i]	A	G	*KIAA0408*	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	*PLCB3*	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	*DNAH10*	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	*CCDC92*	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***	

1712 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
1713 allele; OA=other allele; EAF=effect allele frequency.
1714 a Coding variants refer to variants located in the exons and splicing junction regions.
1715 b Bonferonni corrected Pvalue for the number of SNPs tested for sex-heterogeneity is <7.14E-04 i.e. 0.05/70 variants.
1716 c Variant positions are reported according to Human assembly build 37 and their alleles are coded based on the positive strand.
1717 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).
1718 e Variant was also identified as array-wide significant in the sex-combined analyses.
1719 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).
1720 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
1721 genome-wide association meta-analysis data. Bioinformatics 2015: 31, 259-61. PMID: 25260699.
1722 h Effect size is based on standard deviation (SD) per effect allele
1723 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).
1724 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
1725 of Stage 1 association.
1726