

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

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

24 Kristian Hveem<sup>109,114</sup>, M Arfan Ikram<sup>20,115,116</sup>, Erik Ingelsson<sup>104,117</sup>, Anne U Jackson<sup>18</sup>, Gail P Jarvik<sup>54,118</sup>,  
25 Yucheng Jia<sup>14</sup>, Torben Jørgensen<sup>119,120,121</sup>, Pekka Jousilahti<sup>122</sup>, Johanne M Justesen<sup>50</sup>, Bratati  
26 Kahali<sup>123,124,125,126</sup>, Maria Karaleftheri<sup>127</sup>, Sharon LR Kardia<sup>43</sup>, Fredrik Karpe<sup>128,129</sup>, Frank Kee<sup>130</sup>, Hidetoshi  
27 Kitajima<sup>3</sup>, Pirjo Komulainen<sup>131,132,133</sup>, Jaspal S Kooner<sup>60,62,63,134</sup>, Peter Kovacs<sup>44</sup>, Bernhard K Krämer<sup>135</sup>, Kari  
28 Kuulasmaa<sup>122</sup>, Johanna Kuusisto<sup>136</sup>, Markku Laakso<sup>136</sup>, Timo A Lakka<sup>131,132,133</sup>, David Lamparter<sup>41,42</sup>, Leslie  
29 A Lange<sup>137</sup>, Claudia Langenberg<sup>31</sup>, Eric B Larson<sup>54,138,139</sup>, Nanette R Lee<sup>140,141</sup>, Wen-Jane Lee<sup>142,143</sup>, Terho  
30 Lehtimäki<sup>144,145</sup>, Cora E Lewis<sup>146</sup>, Huaixing Li<sup>111</sup>, Jin Li<sup>147</sup>, Ruifang Li-Gao<sup>77</sup>, Li-An Lin<sup>92</sup>, Xu Lin<sup>111</sup>, Lars  
31 Lind<sup>148</sup>, Jaana Lindström<sup>122</sup>, Allan Linneberg<sup>121,149,150</sup>, Ching-Ti Liu<sup>151</sup>, Dajiang J Liu<sup>152</sup>, Jian'an Luan<sup>31</sup>, Leo-  
32 Pekka Lyytikäinen<sup>144,145</sup>, Stuart MacGregor<sup>153</sup>, Reedik Mägi<sup>22</sup>, Satu Männistö<sup>122</sup>, Gaëlle Marenne<sup>72</sup>,  
33 Jonathan Marten<sup>107</sup>, Nicholas GD Masca<sup>154,155</sup>, Mark I McCarthy<sup>3,128,129</sup>, Karina Meidtner<sup>101,156</sup>, Evelin  
34 Mihailov<sup>22</sup>, Leena Moilanen<sup>157</sup>, Marie Moitry<sup>158,159</sup>, Dennis O Mook-Kanamori<sup>77,160</sup>, Anna Morgan<sup>98</sup>,  
35 Andrew P Morris<sup>3,69</sup>, Martina Müller-Nurasyid<sup>108,161,162</sup>, Patricia B Munroe<sup>16,57</sup>, Narisu Narisu<sup>68</sup>,  
36 Christopher P Nelson<sup>154,155</sup>, Matt Neville<sup>128,129</sup>, Ioanna Ntalla<sup>16</sup>, Jeffrey R O'Connel<sup>163</sup>, Katharine R  
37 Owen<sup>128,129</sup>, Oluf Pedersen<sup>50</sup>, Gina M Peloso<sup>151</sup>, Craig E Pennell<sup>164</sup>, Markus Perola<sup>122,165,166</sup>, James A  
38 Perry<sup>163</sup>, John RB Perry<sup>31</sup>, Tune H Pers<sup>50,167</sup>, Ailith Pirie<sup>86</sup>, Ozren Polasek<sup>168,169</sup>, Olli T Raitakari<sup>170,171</sup>, Asif  
39 Rasheed<sup>172</sup>, Chelsea K Raulerson<sup>137</sup>, Rainer Rauramaa<sup>131,132,133</sup>, Dermot F Reilly<sup>173</sup>, Alex P Reiner<sup>105,174</sup>, Paul  
40 M Ridker<sup>65,66,175</sup>, Manuel A Rivas<sup>11,176</sup>, Neil R Robertson<sup>3,128</sup>, Antonietta Robino<sup>177</sup>, Igor Rudan<sup>169</sup>,  
41 Katherine S Ruth<sup>178</sup>, Danish Saleheen<sup>172,179</sup>, Veikko Salomaa<sup>122</sup>, Nilesh J Samani<sup>154,155</sup>, Pamela J  
42 Schreiner<sup>180</sup>, Matthias B Schulze<sup>101,156</sup>, Robert A Scott<sup>31</sup>, Marcelo P Segura-Lepe<sup>61</sup>, Xueling Sim<sup>18,181</sup>,  
43 Andrew J Slater<sup>182,183</sup>, Kerrin S Small<sup>184</sup>, Blair H Smith<sup>185,186</sup>, Jennifer A Smith<sup>43</sup>, Lorraine Southam<sup>3,72</sup>,  
44 Timothy D Spector<sup>184</sup>, Elizabeth K Speliotes<sup>123,124,125</sup>, Kari Stefansson<sup>187,188</sup>, Valgerdur Steinthorsdottir<sup>187</sup>,  
45 Kathleen E Stirrups<sup>16,33</sup>, Konstantin Strauch<sup>108,189</sup>, Heather M Stringham<sup>18</sup>, Michael Stumvoll<sup>44,45</sup>, Liang  
46 Sun<sup>190,191</sup>, Praveen Surendran<sup>55</sup>, Karin MA Swart<sup>192</sup>, Jean-Claude Tardif<sup>9,85</sup>, Kent D Taylor<sup>14</sup>, Alexander  
47 Teumer<sup>193</sup>, Deborah J Thompson<sup>80</sup>, Gudmar Thorleifsson<sup>187</sup>, Unnur Thorsteinsdottir<sup>187,188</sup>, Betina H

48 Thuesen<sup>121</sup>, Anke Tönjes<sup>194</sup>, Mina Torres<sup>195</sup>, Emmanouil Tsafantakis<sup>196</sup>, Jaakko Tuomilehto<sup>122,197,198,199</sup>,  
49 André G Uitterlinden<sup>20,21</sup>, Matti Uusitupa<sup>200</sup>, Cornelia M van Duijn<sup>20</sup>, Mauno Vanhala<sup>201,202</sup>, Rohit  
50 Varma<sup>195</sup>, Sita H Vermeulen<sup>203</sup>, Henrik Vestergaard<sup>50,204</sup>, Veronique Vitart<sup>107</sup>, Thomas F Vogt<sup>205</sup>, Dragana  
51 Ntalla<sup>99</sup>, Lynne E Wagenknecht<sup>206</sup>, Mark Walker<sup>207</sup>, Lars Wallentin<sup>208</sup>, Feijie Wang<sup>111</sup>, Carol A Wang<sup>164</sup>,  
52 Shuai Wang<sup>151</sup>, Nicholas J Wareham<sup>31</sup>, Helen R Warren<sup>16,57</sup>, Dawn M Waterworth<sup>209</sup>, Jennifer Wessel<sup>210</sup>,  
53 Harvey D White<sup>211</sup>, Cristen J Willer<sup>123,124,212</sup>, James G Wilson<sup>213</sup>, Andrew R Wood<sup>178</sup>, Ying Wu<sup>137</sup>, Hanieh  
54 Yaghoobkar<sup>178</sup>, Jie Yao<sup>14</sup>, Laura M Yerges-Armstrong<sup>163,214</sup>, Robin Young<sup>55,215</sup>, Eleftheria Zeggini<sup>72</sup>, Xiaowei  
55 Zhan<sup>216</sup>, Weihua Zhang<sup>60,61</sup>, Jing Hua Zhao<sup>31</sup>, Wei Zhao<sup>179</sup>, He Zheng<sup>111</sup>, Wei Zhou<sup>123,124</sup>, M Carola  
56 Zillikens<sup>20,21</sup>, CHD Exome+ Consortium, EPIC-CVD Consortium, ExomeBP Consortium, Global Lipids  
57 Genetic Consortium, GoT2D Genes Consortium, InterAct, ReproGen Consortium, T2D-Genes  
58 Consortium, The MAGIC Investigators, Fernando Rivadeneira<sup>20,21</sup>, Ingrid B Borecki<sup>28</sup>, John A Pospisilik<sup>15</sup>,  
59 Panos Deloukas<sup>16,217</sup>, Timothy M Frayling<sup>178</sup>, Guillaume Lettre<sup>9,85</sup>, Karen L Mohlke<sup>137</sup>, Jerome I Rotter<sup>14</sup>,  
60 Zoltán Kutalik<sup>42,218</sup>, Joel N Hirschhorn<sup>11,13,219</sup>, L Adrienne Cupples<sup>□,27,151</sup>, Ruth JF Loos<sup>□,7,8,220</sup>, \*Kari E  
61 North<sup>\*,□,221</sup>, \*Cecilia M Lindgren<sup>\*,□,3,11,222</sup>

62 ¥ These authors contributed equally to this work.

63 □ These authors jointly supervised this work.

#### 64 \*CORRESPONDING AUTHORS

65 Prof. Kari North

66 Department of Epidemiology

67 University of North Carolina at Chapel Hill

68 137 East Franklin Street

69 Suite 306

70 Chapel Hill, NC 27514

71

72 Prof. Cecilia M Lindgren

73 The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery

74 University of Oxford  
75 Roosevelt Drive  
76 Oxford  
77 OX3 7BN  
78 United Kingdom  
79 [celi@well.ox.ac.uk](mailto:celi@well.ox.ac.uk)

## 80 **AFFILIATIONS**

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

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

- 122 28. Division of Statistical Genomics, Department of Genetics, Washington University School of  
123 Medicine, St. Louis, MO, 63108, USA
- 124 29. Department of Medicine, Harvard University Medical School, Boston, MA, 02115, USA
- 125 30. Massachusetts General Hospital, Boston, MA, 02114, USA
- 126 31. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of  
127 Metabolic Science, Cambridge, CB2 0QQ, UK
- 128 32. Department of Vascular Medicine, AMC, Amsterdam, 1105 AZ, The Netherlands
- 129 33. Department of Haematology, University of Cambridge, Cambridge, CB2 0PT, UK
- 130 34. School of Kinesiology and Health Science, Faculty of Health, York University, Toronto
- 131 35. Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA,  
132 92093, USA
- 133 36. INSERM U1167, Lille, F-59019, France
- 134 37. Institut Pasteur de Lille, U1167, Lille, F-59019, France
- 135 38. Universite de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related  
136 diseases, Lille, F-59019, France
- 137 39. INSERM U1018, Centre de recherche en Épidémiologie et Sante des Populations (CESP), Villejuif,  
138 France
- 139 40. Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, 37203, USA
- 140 41. Department of Computational Biology, University of Lausanne, Lausanne, 1011, Switzerland
- 141 42. Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
- 142 43. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI,  
143 48109, USA
- 144 44. IFB Adiposity Diseases, University of Leipzig, Leipzig, 04103, Germany
- 145 45. University of Leipzig, Department of Medicine, Leipzig, 04103, Germany

- 146 46. Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE),  
147 Nuthetal, 14558, Germany
- 148 47. School of Public Health, Human Genetics Center, The University of Texas Health Science Center  
149 at Houston, Houston, TX, 77030, USA
- 150 48. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030 USA
- 151 49. Department of Nephrology, University Hospital Regensburg, Regensburg, 93042, Germany
- 152 50. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and  
153 Medical Sciences, University of Copenhagen, Copenhagen, 2100, Denmark
- 154 51. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
- 155 52. Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, 7100, Denmark
- 156 53. Institute of Regional Health Research, University of Southern Denmark, Odense, 5000, Denmark
- 157 54. Department of Medicine, University of Washington, Seattle, WA, 98195, USA
- 158 55. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care,  
159 University of Cambridge, Cambridge, CB1 8RN, UK
- 160 56. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public  
161 Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
- 162 57. NIHR Barts Cardiovascular Research Unit, Barts and The London School of Medicine & Dentistry,  
163 Queen Mary University of London, London, EC1M 6BQ, UK
- 164 58. Research Centre on Public Health, University of Milano-Bicocca, Monza, 20900, Italy
- 165 59. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232,  
166 Singapore
- 167 60. Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital,  
168 Middlesex, UB1 3HW, UK

- 169 61. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,  
170 London, W2 1PG, UK
- 171 62. Imperial College Healthcare NHS Trust, London, W12 0HS, UK
- 172 63. MRC-PHE Centre for Environment and Health, Imperial College London, London, W2 1PG, UK
- 173 64. Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA,  
174 02115, USA
- 175 65. Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston,  
176 MA, 02215, USA
- 177 66. Harvard Medical School, Boston, MA, 02115, USA
- 178 67. Medical department, Lillebaelt Hospital, Vejle, 7100, Denmark
- 179 68. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute,  
180 National Institutes of Health, Bethesda, MD, 20892, USA
- 181 69. Department of Biostatistics, University of Liverpool, Liverpool, L69 3GL, UK
- 182 70. Menzies Health Institute Queensland, Griffith University, Southport, QLD, Australia
- 183 71. Department of Biomedical Informatics and Medical Education, University of Washington, Seattle,  
184 WA, 98195, USA
- 185 72. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK
- 186 73. British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of  
187 Cambridge, Cambridge, CB2 0QQ, UK
- 188 74. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht,  
189 The Netherlands
- 190 75. Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht,  
191 Utrecht, 3584 CX, The Netherlands
- 192 76. Faculty of Pharmacy, Universite de Montreal, Montreal, Quebec, H3T 1J4, Canada



- 193 77. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2300RC, The  
194 Netherlands
- 195 78. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio  
196 University, Athens, 17671, Greece
- 197 79. Division of Epidemiology & Community Health, School of Public Health, University of Minnesota,  
198 Minneapolis, MN, 55454, USA
- 199 80. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care,  
200 University of Cambridge, Cambridge, CB1 8RN, UK
- 201 81. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, 17475,  
202 Germany
- 203 82. DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, 17475,  
204 Germany
- 205 83. Institute of Cardiovascular Science, University College London, London, WC1E 6JF, UK
- 206 84. MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of  
207 Bristol, Bristol, BS8 2BN, UK
- 208 85. Department of Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, H3T  
209 1J4, Canada
- 210 86. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge,  
211 Cambridge, CB1 8RN, UK
- 212 87. Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health,  
213 School of Public Health, Imperial College London, London, W2 1PG, UK
- 214 88. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina,  
215 45110, Greece
- 216 89. CNR Institute of Clinical Physiology, Pisa, Italy

- 217 90. Department of Clinical & Experimental Medicine, University of Pisa, Italy
- 218 91. Toulouse University School of Medicine, Toulouse, TSA 50032 31059, France
- 219 92. Institute of Molecular Medicine, The University of Texas Health Science Center at Houston,  
220 Houston, TX, 77030, USA
- 221 93. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University,  
222 Malmo, SE-20502, Sweden
- 223 94. Department of Nutrition, Harvard School of Public Health, Boston, MA, 02115, USA
- 224 95. Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå,  
225 901 87, Sweden
- 226 96. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald,  
227 Greifswald, 17475, Germany
- 228 97. Ilaria Gandin, Research Unit, AREA Science Park, Trieste, 34149, Italy
- 229 98. Department of Medical Sciences, University of Trieste, Trieste, 34137, Italy
- 230 99. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy
- 231 100. Geriatrics, Department of Public Health, Uppsala University, Uppsala, 751 85, Sweden
- 232 101. German Center for Diabetes Research, München-Neuherberg, 85764, Germany  
233
- 234 102. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for  
235 Environmental Health, Neuherberg, 85764, Germany
- 236 103. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research  
237 Center for Environmental Health, Neuherberg, 85764, Germany
- 238 104. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory,  
239 Uppsala University, Uppsala, 751 41, Sweden

- 240 105. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, 98109,  
241 USA
- 242 106. University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
- 243 107. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of  
244 Edinburgh, Edinburgh, EH4 2XU, UK
- 245 108. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for  
246 Environmental Health, Neuherberg, 85764, Germany
- 247 109. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian  
248 University of Science and Technology, Trondheim, 7600, Norway
- 249 110. AMC, Department of Vascular Medicine, Amsterdam, 1105 AZ, The Netherlands
- 250 111. Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai  
251 Institutes for Biological Sciences, Chinese Academy of Sciences, University of the Chinese  
252 Academy of Sciences, Shanghai, People's Republic of China, Shanghai, 200031, China
- 253 112. Division of Endocrinology and Metabolism, Department of Internal Medicine, Tri-Service General  
254 Hospital, Taipei, Taiwan 114, Taiwan
- 255 113. School of Medicine, National Defense Medical Center, Taipei, Taiwan 114, Taiwan
- 256 114. HUNT Research center, Department of Public Health, Norwegian University of Science and  
257 Technology, Levanger, 7600, Norway
- 258 115. Department of Neurology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- 259 116. Department of Radiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- 260 117. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of  
261 Medicine, Stanford, CA, 943 05, USA
- 262 118. Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA
- 263 119. Faculty of medicine, Aalborg University, Aalborg, DK-9000, Denmark

- 264 120. Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen,  
265 Copenhagen, 2200, Denmark
- 266 121. Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, DK-2600,  
267 Denmark
- 268 122. National Institute for Health and Welfare, Helsinki, FI-00271, Finland
- 269 123. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor,  
270 MI, 48109, USA
- 271 124. Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109, USA
- 272 125. Division of Gastroenterology, University of Michigan, Ann Arbor, MI, 48109, USA
- 273 126. Centre for Brain Research, Indian Institute of Science, Bangalore 560012, India
- 274 127. Echinus Medical Centre, Echinus, Greece
- 275 128. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine,  
276 University of Oxford, Oxford, OX3 7LE, UK
- 277 129. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE,  
278 UK
- 279 130. UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast, UK,  
280 BT12 6BJ, UK
- 281 131. Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise  
282 Medicine, Kuopio, 70100, Finland
- 283 132. Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus,  
284 70210, Finland
- 285 133. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio,  
286 Finland

- 287 134. National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus,  
288 London, W12 0NN, UK
- 289 135. University Medical Centre Mannheim, 5th Medical Department, University of Heidelberg,  
290 Mannheim, 68167, Germany
- 291 136. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio  
292 University Hospital, Kuopio, 70210, Finland
- 293 137. Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA
- 294 138. Kaiser Permanente Washington Health Research Institute Seattle WA 98101
- 295 139. Department of Health Services, University of Washington, Seattle WA 98101
- 296 140. Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, 6000,  
297 Philippines
- 298 141. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, 6000,  
299 Philippines
- 300 142. Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan 407,  
301 Taiwan
- 302 143. Department of Social Work, Tunghai University, Taichung, Taiwan
- 303 144. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, 33521, Finland
- 304 145. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of  
305 Medicine and Life Sciences, University of Tampere, Tampere 33014, Finland
- 306 146. Division of Preventive Medicine University of Alabama at Birmingham, Birmingham, AL 35205,  
307 USA
- 308 147. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of  
309 Medicine, Palo Alto, CA, 94304, USA
- 310 148. Uppsala University, Uppsala, 75185, Sweden

- 311 149. Department of Clinical Experimental Research, Rigshospitalet, Copenhagen, DK-2200, Denmark
- 312 150. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of  
313 Copenhagen, Copenhagen, 2200, Denmark
- 314 151. Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA
- 315 152. Department of Public Health Sciences, Institute for Personalized Medicine, the Pennsylvania  
316 State University College of Medicine, Hershey, PA, 17033, USA
- 317 153. QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia
- 318 154. Department of Cardiovascular Sciences, Univeristy of Leicester, Glenfield Hospital, Leicester, LE3  
319 9QP, UK
- 320 155. NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP,  
321 UK
- 322 156. Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-  
323 Rehbruecke (DIfE), Nuthetal, 14558, Germany
- 324 157. Department of Medicine, Kuopio University Hospital, Kuopio, 70210, Finland
- 325 158. Department of Epidemiology and Public Health, University of Strasbourg, Strasbourg, F-67085,  
326 France
- 327 159. Department of Public Health, University Hospital of Strasbourg, Strasbourg, F-67081, France
- 328 160. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden,  
329 2300RC, The Netherlands
- 330 161. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universitat,  
331 Munich, 81377, Germany
- 332 162. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich,  
333 80802, Germany

- 334 163. Program for Personalized and Genomic Medicine, Department of Medicine, University of  
335 Maryland School of Medicine, Baltimore, MD, 21201, US
- 336 164. School of Women's and Infants' Health, The University of Western Australia, Perth, Western  
337 Australia, 6009, Australia
- 338 165. University of Helsinki, Institute for Molecular Medicine (FIMM) and Diabetes and Obesity  
339 Research Program, Helsinki, FI00014, Finland
- 340 166. University of Tartu, Estonian Genome Center, Tartu, Estonia, Tartu, 51010, Estonia
- 341 167. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, 2200, Denmark
- 342 168. School of Medicine, University of Split, Split, 21000, Croatia
- 343 169. Centre for Global Health Research, Usher Institute of Population Health Sciences and  
344 Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK
- 345 170. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku,  
346 20521, Finland
- 347 171. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku,  
348 20520, Finland
- 349 172. Centre for Non-Communicable Diseases, Karachi, Pakistan
- 350 173. Merck, Sharp & Dohme, Genetics and Pharmacogenomics, Boston, MA, 02115, USA
- 351 174. Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA
- 352 175. Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical  
353 School, Boston, MA, 02115, USA
- 354 176. Nuffield Department of Clinical Medicine, Oxford, OX37 BN, UK
- 355 177. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, 34137, Italy
- 356 178. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2  
357 5DW, UK

- 358 179. Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of  
359 Pennsylvania, Philadelphia, PA, 19104, USA
- 360 180. Division of Epidemiology & Community Health University of Minnesota, Minneapolis, MN,  
361 55454, USA
- 362 181. Saw Swee Hock School of Public Health, National University Health System, National University  
363 of Singapore, Singapore 117549, Singapore
- 364 182. Genetics, Target Sciences, GlaxoSmithKline, Research Triangle Park, NC, 27709, US
- 365 183. OmicSoft a QIAGEN Company, Cary, NC, 27513, US
- 366 184. Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1  
367 7EH, UK
- 368 185. Division of Population Health Sciences, Ninewells Hospital and Medical School, University of  
369 Dundee, Dundee, UK
- 370 186. Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh,  
371 Edinburgh, EH4 2XU, UK
- 372 187. deCODE Genetics/Amgen inc., Reykjavik, 101, Iceland
- 373 188. Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland
- 374 189. Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, 81377, Germany
- 375 190. Biodemography of Aging Research Unit, Social Science Research Institute, Duke University,  
376 Durham, NC, 27708, USA
- 377 191. Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland
- 378 192. VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam, 1007  
379 MB, The Netherlands
- 380 193. Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany



- 381 194. Center for Pediatric Research, Department for Women's and Child Health, University of Leipzig,  
382 Leipzig, 04103, Germany
- 383 195. USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the  
384 University of Southern California, Los Angeles, CA, 90033, USA
- 385 196. Anogia Medical Centre, Anogia, Greece
- 386 197. Centre for Vascular Prevention, Danube-University Krems, Krems, 3500, Austria
- 387 198. Dasman Diabetes Institute, Dasman, 15462, Kuwait
- 388 199. Diabetes Research Group, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
- 389 200. Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, 70210,  
390 Finland
- 391 201. Central Finland Central Hospital, Jyvaskyla, 40620, Finland
- 392 202. University of Eastern Finland, Kuopio, 70210, Finland
- 393 203. Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, 6500 HB,  
394 The Netherlands
- 395 204. Steno Diabetes Center Copenhagen, Gentofte, 2800, Denmark
- 396 205. Merck, Sharp & Dohme, Cardiometabolic Disease, Kenilworth, NJ, 07033, USA
- 397 206. Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, 27157,  
398 USA
- 399 207. Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, NE2 4HH,  
400 UK
- 401 208. Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala  
402 University, Uppsala, 752 37, Sweden
- 403 209. Genetics, Target Sciences, GlaxoSmithKline, King of Prussia, PA, US

- 404 210. Departments of Epidemiology & Medicine, Diabetes Translational Research Center, Fairbanks  
405 School of Public Health & School of Medicine, Indiana University, Indiana, IN, 46202, USA
- 406 211. Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland,  
407 New Zealand
- 408 212. Department of Human Genetics, University of Michigan, Ann Arbor, MI, 48109, USA
- 409 213. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS,  
410 39216, USA
- 411 214. GlaxoSmithKline, King of Prussia, PA, 19406, USA
- 412 215. University of Glasgow, Glasgow, G12 8QQ, UK
- 413 216. Department of Clinical Sciences, Quantitative Biomedical Research Center, Center for the  
414 Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, 75390,  
415 USA
- 416 217. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-  
417 HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia
- 418 218. Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, 1010,  
419 Switzerland
- 420 219. Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, 02115, USA
- 421 220. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai,  
422 New York, NY, 10069, USA
- 423 221. Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, NC, 27514,  
424 USA
- 425 222. Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of  
426 Oxford, Oxford, OX3 7BN, UK
- 427

428 **ABSTRACT**

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

444  
445  
446  
447  
448  
449

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

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

## 467 **RESULTS**

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

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

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

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

496 effects, our results point to the importance of investigating rare and low frequency variants to identify  
497 variants with large effects (**Figure 2**).

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

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

520 effect estimate on average). For 4 of the 55 SNPs (rs141845046, rs1034405, rs3617, rs9469913, **Table 1**),  
521 the association with WHRadjBMI appears to be attenuated following correction ( $P_{\text{corrected}} > 9 \times 10^{-4}$ ,  
522 0.05/55), including one novel variant, rs1034405 in *C3orf18*. Thus, these 4 variants warrant further  
523 functional investigations to quantify their impact on WHR, as a true association may still exist, although  
524 the effect may be slightly overestimated in the current analysis.

525 Using stage 1 meta-analysis results, we then aggregated low frequency variants across genes  
526 and tested their joint effect with both SKAT and burden tests<sup>20</sup> (**Supplementary Table 8, Online**  
527 **Methods**). We identified five genes that reached array-wide significance ( $P < 2.5 \times 10^{-6}$ , 0.05/16,222 genes  
528 tested), *RAPGEF3*, *ACVR1C*, *ANGPTL4*, *DNAI1*, and *NOP2*. However, while all genes analyzed included  
529 more than one variant, none remained significant after conditioning on the single variant with the most  
530 significant p-value. We identified variants within *RAPGEF3*, *ACVR1C*, *ANGPTL4* that reached suggestive  
531 significance in Stage 1 and chip-wide significance in stage 1+2 for one or more meta-analyses (**Tables 1**  
532 **and 2**); however, we did not identify any significant variants for *DNAI1* and *NOP2*. While neither of these  
533 genes had a single variant that reached chip-wide significance, they each had variants with nearly  
534 significant results (*NOP2*:  $P = 3.69 \times 10^{-5}$ , *DNAI1*:  $4.64 \times 10^{-5}$ ). Combined effects with these single variants  
535 and others in LD within the gene likely drove the association in our aggregate gene-based tests, but  
536 resulted in non-significance following conditioning on the top variant. While our results suggest these  
537 associations are driven by a single variant, each gene may warrant consideration in future investigations.

538

### 539 **Conditional analyses**

540 We next implemented conditional analyses to determine (1) the number of independent  
541 association signals the 56 array-wide significant coding variants represent, and (2) whether the 33  
542 variants near known GWAS association signals ( $\pm 1$ Mb) represent independent novel association  
543 signals. To determine if these variants were independent association signals, we used approximate joint

544 conditional analyses to test for independence in stage 1 (**Online Methods; Supplementary Table 4**)<sup>20</sup>.  
545 Only the *RSPO3-KIAA0408* locus contains two independent variants 291 Kb apart, rs1892172 in *RSPO3*  
546 (MAF=46.1%,  $P_{\text{conditional}}=4.37 \times 10^{-23}$  in the combined sexes, and  $P_{\text{conditional}}=2.4 \times 10^{-20}$  in women) and  
547 rs139745911 in *KIAA0408* (MAF=0.9%,  $P_{\text{conditional}}=3.68 \times 10^{-11}$  in the combined sexes, and  
548  $P_{\text{conditional}}=1.46 \times 10^{-11}$  in women; **Figure 3A**).

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

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

## 562 **Gene set and pathway enrichment analysis**

563 To determine if the significant coding variants highlight novel biological pathways and/or  
564 provide additional support for previously identified biological pathways, we applied two complementary  
565 pathway analysis methods using the EC-DEPICT (ExomeChip Data-driven Expression Prioritized  
566 Integration for Complex Traits) pathway analysis tool,<sup>21,22</sup> and PASCAL<sup>23</sup> (**Online Methods**). While for



567 PASCAL all variants were used, in the case of EC-DEPICT, we examined 361 variants with suggestive  
568 significance ( $P < 5 \times 10^{-4}$ )<sup>10,17</sup> from the combined ancestries and combined sexes analysis (which after  
569 clumping and filtering became 101 lead variants in 101 genes). We separately analyzed variants that  
570 exhibited significant sex-specific effects ( $P_{\text{sexhet}} < 5 \times 10^{-4}$ ).

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

584 To focus on novel findings, we conducted pathway analyses after excluding variants from  
585 previous WHRadjBMI analyses<sup>10</sup> (**Supplemental Note 2**). Seventy-five loci/genes were included in the  
586 EC-DEPICT analysis, and we identified 26 significantly enriched gene sets (13 meta-gene sets). Here, all  
587 but one gene set, “lipid particle size”, were related to skeletal biology. This result likely reflects an effect  
588 on the pelvic skeleton (hip circumference), shared signaling pathways between bone and fat (such as  
589 TGF-beta) and shared developmental origin<sup>24</sup> (**Supplementary Data 5, Supplementary Figure 16b**).

590 Many of these pathways were previously found to be significant in the GWAS DEPICT analysis; these  
591 findings provide a fully independent replication of their biological relevance for WHRadjBMI.

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

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

## 611 **Cross-trait associations**

612 To assess the relevance of our identified variants with cardiometabolic, anthropometric, and  
613 reproductive traits, we conducted association lookups from existing ExomeChip studies of 15 traits  
614 (**Supplementary Data 6, Supplementary Figure 19**). Indeed, the clinical relevance of central adiposity is  
615 likely to be found in the cascade of impacts such variants have on downstream cardiometabolic  
616 disease.<sup>22,25-29</sup> We found that variants in *STAB1* and *PLCB3* display the greatest number of significant  
617 cross-trait associations, each associating with seven different traits ( $P < 9.8 \times 10^{-4}$ , 0.05/51 variants tested).  
618 Of note, these two genes cluster together with *RSPO3*, *DNAH10*, *MNS1*, *COBLL1*, *CCDC92*, and *ITIH3*  
619 (**Supplementary Data 6, Supplementary Figure 19**). The WHR-increasing alleles in this cluster of variants  
620 exhibit a pattern of increased cardiometabolic risk (e.g. increased fasting insulin [FI], two-hour glucose  
621 [TwoHGlu], and triglycerides [TG]; and decreased high-density lipoprotein cholesterol [HDL]), but also  
622 decreased BMI. This phenomenon, where variants associated with lower BMI are also associated with  
623 increased cardiometabolic risk, has been previously reported.<sup>30-36</sup> A recent Mendelian Randomization  
624 (MR) analysis of the relationship between central adiposity (measured as WHRadjBMI) and  
625 cardiometabolic risk factors found central adiposity to be causal.<sup>9</sup> Using 48 WHR-increasing variants  
626 reported in the recent GIANT analysis<sup>10</sup> to calculate a polygenic risk score, Emdin *et al.* found that a 1 SD  
627 increase in genetic risk of central adiposity was associated with higher total cholesterol, triglyceride  
628 levels, fasting insulin and two-hour glucose, and lower HDL – all indicators of cardiometabolic disease,  
629 and also associated with a 1 unit decrease in BMI<sup>9</sup>.

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

636 our ExomeChip variants are in LD with GWAS variants associated with other behavioral and neurological  
637 traits (schizophrenia, bipolar disorder *etc.*), and inflammatory or autoimmune diseases (Crohn's Disease,  
638 multiple sclerosis *etc.*) (**Supplementary Data 7**).

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

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

## 657 **Genetic architecture of WHRadjBMI coding variants**

658 We used summary statistics from our stage 1 results to estimate the phenotypic variance  
659 explained by ExomeChip coding variants. We calculated the variance explained by subsets of SNPs across  
660 various significance thresholds ( $P < 2 \times 10^{-7}$  to 0.2) and conservatively estimated using only independent  
661 tag SNPs (**Supplementary Table 12, Online Methods, and Supplementary Figure 22**). The 22  
662 independent significant coding SNPs in stage 1 account for 0.28% of phenotypic variance in WHRadjBMI.  
663 For independent variants that reached suggestive significance in stage 1 ( $P < 2 \times 10^{-6}$ ), 33 SNPs explain  
664 0.38% of the variation; however, the 1,786 independent SNPs with a liberal threshold of  $P < 0.02$  explain  
665 13 times more variation (5.12%). While these large effect estimates may be subject to winner's curse,  
666 for array-wide significant variants, we detected a consistent relationship between effect magnitude and  
667 MAF in our stage 2 analyses in UK Biobank and deCODE (**Supplementary Data 1-3**). Notably, the  
668 Exomechip coding variants explained less of the phenotypic variance than in our previous GIANT  
669 investigation, wherein 49 significant SNPs explained 1.4% of the variance in WHRadjBMI. When  
670 considering all coding variants on the ExomeChip in men and women together, 46 SNPs with a  $P < 2 \times 10^{-6}$   
671 and 5,917 SNPs with a  $P < 0.02$  explain 0.51% and 13.75% of the variance in WHRadjBMI, respectively. As  
672 expected given the design of the ExomeChip, the majority of the variance explained is attributable to  
673 rare and low frequency coding variants (independent SNPs with  $MAF < 1\%$  and  $MAF < 5\%$  explain 5.18%  
674 and 5.58%, respectively). However, for rare and low frequency variants, those that passed significance in  
675 stage 1 explain only 0.10% of the variance in WHRadjBMI. As in **Figure 2**, these results also indicate that  
676 there are additional coding variants associated with WHRadjBMI that remain to be discovered,  
677 particularly rare and low frequency variants with larger effects than common variants. Due to observed  
678 differences in association strength between women and men, we estimated variance explained for the  
679 same set of SNPs in women and men separately. As observed in previous studies<sup>10</sup>, there was  
680 significantly ( $P_{RsqDiff} < 0.002 = 0.05/21$ , Bonferroni-corrected threshold) more variance explained in women  
681 compared to men at each significance threshold considered (differences ranged from 0.24% to 0.91%).

682 To better understand the potential clinical impact of WHRadjBMI associated variants, we  
683 conducted penetrance analysis using the UKBB population (both sexes combined, and men- and women-  
684 only). We compared the number of carriers and non-carriers of the minor allele for each of our  
685 significant variants in centrally obese and non-obese individuals to determine if there is a significant  
686 accumulation of the minor allele in either the centrally obese or non-obese groups (**Online Methods**).  
687 Three rare and low frequency variants ( $MAF \leq 1\%$ ) with larger effect sizes (effect size  $> 0.90$ ) were  
688 included in the penetrance analysis using World Health Organization (WHO- obese women  $WHR > 0.85$   
689 and obese men  $WHR > 0.90$ ) WHR cut-offs for central obesity. Of these, one SNV (rs55920843-ACVR1C;  
690  $P_{sex-combined} = 9.25 \times 10^{-5}$ ;  $P_{women} = 4.85 \times 10^{-5}$ ) showed a statistically significant difference in the number of  
691 carriers and non-carriers of the minor allele when the two strata were compared (sex-combined obese  
692 carriers=2.2%; non-obese carriers=2.6%; women obese carriers=2.1%; non-obese women carriers=2.6%  
693 (**Supplementary Table 13, Supplementary Figure 23**). These differences were significant in women, but  
694 not in men ( $P_{men} < 5.5 \times 10^{-3}$  after Bonferroni correction for 9 tests) and agree with our overall meta-  
695 analysis results, where the minor allele (G) was significantly associated with lower WHRadjBMI in  
696 women only (**Tables 1 and 2**).

## 697 **Evidence for functional role of significant variants**

### 698 ***Drosophila* Knockdown**

699 Considering the genetic evidence of adipose and insulin biology in determining body fat  
700 distribution<sup>10</sup>, and the lipid signature of the variants described here, we examined whole-body  
701 triglycerides levels in adult *Drosophila*, a model organism in which the fat body is an organ functionally  
702 analogous to mammalian liver and adipose tissue and triglycerides are the major source of fat storage<sup>45</sup>.  
703 Of the 51 genes harboring our 56 significantly associated variants, we identified 27 with *Drosophila*  
704 orthologues for functional follow-up analyses. In order to prioritize genes for follow-up, we selected  
705 genes with large changes in triglyceride storage levels ( $> 20\%$  increase or  $> 40\%$  decrease, as chance

706 alone is unlikely to cause changes of this magnitude, although some decrease is expected) after  
707 considering each corresponding orthologue in an existing large-scale screen for adipose with  $\leq 2$   
708 replicates per knockdown strain.<sup>45</sup> Two orthologues, for *PLXND1* and *DNAH10*, from two separate loci  
709 met these criteria. For these two genes, we conducted additional knockdown experiments with  $\geq 5$   
710 replicates using tissue-specific drivers (fat body [cg-Gal4] and neuronal [elav-Gal4] specific RNAi-  
711 knockdowns) (**Supplementary Table 14**). A significant ( $P < 0.025$ , 0.05/2 orthologues) increase in the total  
712 body triglyceride levels was observed in *DNAH10* orthologue knockdown strains for both the fat body  
713 and neuronal drivers. However, only the neuronal driver knockdown for *PLXND1* produced a significant  
714 change in triglyceride storage. *DNAH10* and *PLXND1* both lie within previous GWAS identified regions.  
715 Adjacent genes have been highlighted as likely candidates for the *DNAH10* association region, including  
716 *CCDC92* and *ZNF664* based on eQTL evidence. However, our fly knockdown results support *DNAH10* as  
717 the causal genes underlying this association. Of note, rs11057353 in *DNAH10* showed suggestive  
718 significance after conditioning on the known GWAS variants in nearby *CCDC92* (sex-combined  
719  $P_{\text{conditional}} = 7.56 \times 10^{-7}$ ; women-only rs11057353  $P_{\text{conditional}} = 5.86 \times 10^{-7}$ , **Supplementary Table 6**; thus  
720 providing some evidence of multiple causal variants/genes underlying this association signal. Further  
721 analyses are needed to determine whether the implicated coding variants from the current analysis are  
722 the putatively functional variants, specifically how these variants affect transcription in and around  
723 these loci, and exactly how those effects alter biology of relevant human metabolic tissues.

#### 724 ***eQTL Lookups***

725 To gain a better understanding of the potential functionality of novel and low frequency  
726 variants, we examined the *cis*-association of the identified variants with expression level of nearby genes  
727 in subcutaneous adipose tissue, visceral omental adipose tissue, skeletal muscle and pancreas from  
728 GTEX<sup>46</sup>, and assessed whether the exome and eQTL associations implicated the same signal (**Online**

729 **Methods, Supplementary Data 9, Supplementary Table 15**). The lead exome variant was associated  
730 with expression level of the coding gene itself for *DAGLB*, *MLXIPL*, *CCDC92*, *MAPKBP1*, *LRRC36* and  
731 *UQCC1*. However, at three of these loci (*MLXIPL*, *MAPKBP1*, and *LRRC36*), the lead exome variant is also  
732 associated with expression level of additional nearby genes, and at three additional loci, the lead exome  
733 variant is only associated with expression level of nearby genes (*HEMK1* at *C3orf18*; *NT5DC2*, *SMIM4*  
734 and *TMEM110* at *STAB1/ITIH3*; and *C6orf106* at *UHRF1BP1*). Although detected with a missense variant,  
735 these loci are also consistent with a regulatory mechanism of effect as they are significantly associated  
736 with expression levels of genes, and the association signal may well be due to LD with nearby regulatory  
737 variants.

738         Some of the coding genes implicated by eQTL analyses are known to be involved in adipocyte  
739 differentiation or insulin sensitivity: e. g. for *MLXIPL*, the encoded carbohydrate responsive element  
740 binding protein is a transcription factor, regulating glucose-mediated induction of *de novo* lipogenesis in  
741 adipose tissue, and expression of its *beta*-isoform in adipose tissue is positively correlated with adipose  
742 insulin sensitivity<sup>47,48</sup>. For *CCDC92*, the reduced adipocyte lipid accumulation upon knockdown  
743 confirmed the involvement of its encoded protein in adipose differentiation<sup>49</sup>.

#### 744 **Biological Curation**

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



752

## 753 **DISCUSSION**

754 Our two-staged approach to analysis of coding variants from ExomeChip data in up to 476,546  
755 individuals identified a total of 56 array-wide significant variants in 41 independent association signals,  
756 including 24 newly identified (23 novel and one independent of known GWAS signals) that influence  
757 WHRadjBMI. Nine of these variants were low frequency or rare, indicating an important role for low  
758 frequency variants in the polygenic architecture of fat distribution and providing further insights into its  
759 underlying etiology. While, due to their rarity, these coding variants only explain a small proportion of  
760 the trait variance at a population level, they may, given their predicted role, be more functionally  
761 tractable than non-coding variants and have a critical impact at the individual and clinical level. For  
762 instance, the association between a low frequency variant (rs11209026; R381Q; MAF<5% in ExAC)  
763 located in the *IL23R* gene and multiple inflammatory diseases (such as psoriasis<sup>50</sup>, rheumatoid arthritis<sup>51</sup>,  
764 ankylosing spondylitis<sup>52</sup>, and inflammatory bowel diseases<sup>53</sup>) led to the development of new therapies,  
765 targeting *IL23* and *IL12* in the same pathway (reviewed in <sup>54-56</sup>). Thus, we are encouraged that our  
766 associated low frequency coding variants displayed large effect sizes; all but one of the nine novel low  
767 frequency variants had an effect size larger than the 49 SNPs reported in Shungin *et al.* 2015, and some  
768 of these effect sizes were up to 7-fold larger than those previously reported for GWAS. This finding  
769 mirrors results for other cardiometabolic traits<sup>57</sup>, and suggests variants of possible clinical significance  
770 with even larger effect and lower frequency variants will likely be detected through larger additional  
771 genome-wide scans of many more individuals.

772 We continue to observe sexual dimorphism in the genetic architecture of WHRadjBMI<sup>11</sup>. Overall,  
773 we identified 19 coding variants that display significant sex differences, of which 16 (84%) display larger  
774 effects in women compared to men. Of the variants outside of GWAS loci, we reported three (two with

775 MAF<5%) that show a significantly stronger effect in women and two (one with MAF<5%) that show a  
776 stronger effect in men. Additionally, genetic variants continue to explain a higher proportion of the  
777 phenotypic variation in body fat distribution in women compared to men<sup>10,11</sup>. Of the novel female  
778 (*DSTYK* and *ANGPTL4*) and male (*UGGT2* and *MMP14*) specific signals, only *ANGPTL4* implicated fat  
779 distribution related biology associated with both lipid biology and cardiovascular traits (**Box 1**). Sexual  
780 dimorphism in fat distribution is apparent from childhood and throughout adult life<sup>58-60</sup>, and at sexually  
781 dimorphic loci, hormones with different levels in men and women may interact with genomic and  
782 epigenomic factors to regulate gene activity, though this remains to be experimentally documented.  
783 Dissecting the underlying molecular mechanisms of the sexual dimorphism in body fat distribution, and  
784 also how it is correlated with – and causing – important comorbidities like T2D and cardiovascular  
785 diseases will be crucial for improved understanding of disease risk and pathogenesis.

786 Overall, we observe fewer significant associations between WHRadjBMI and coding variants on  
787 the ExomeChip than Turcot *et al.*<sup>25</sup> examining the association of low frequency and rare coding variants  
788 with BMI. In line with these observations, we identify fewer pathways and cross-trait associations. One  
789 reason for fewer WHRadjBMI implicated variants and pathways may be smaller sample size ( $N_{\text{WHRadjBMI}} =$   
790 476,546,  $N_{\text{BMI}} = 718,639$ ), and thus, lower statistical power. Power, however, is likely not the only  
791 contributing factor. For example, Turcot *et al.*<sup>25</sup> have comparative sample sizes between BMI and that  
792 of Marouli *et al.*<sup>22</sup> studying height ( $N_{\text{height}} = 711,428$ ). However, greater than seven times the number of  
793 coding variants are identified for height than for BMI, indicating that perhaps a number of other factors,  
794 including trait architecture, heritability (possibly overestimated in some phenotypes), and phenotype  
795 precision, likely all contribute to our study's capacity to identify low frequency and rare variants with  
796 large effects. Further, it is possible that the comparative lack of significant findings for WHRadjBMI and  
797 BMI compared to height may be a result of higher selective pressure against genetic predisposition to  
798 cardiometabolic phenotypes, such as BMI and WHR. As evolutionary theory predicts that harmful alleles

799 will be low frequency<sup>61</sup>, we may need larger sample sizes to detect rare variants that have so far  
800 escaped selective pressures. Lastly, the ExomeChip is limited by the variants that are present on the  
801 chip, which was largely dictated by sequencing studies in European-ancestry populations and a MAF  
802 detection criteria of ~0.012%. It is likely that through an increased sample size, use of chips designed to  
803 detect variation across a range of continental ancestries, high quality, deep imputation with large  
804 reference samples (e.g. HRC), and/or alternative study designs, future studies will detect additional  
805 variation from the entire allele frequency spectrum that contributes to fat distribution phenotypes.

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

819         The ultimate aim of genetic investigations of obesity-related traits, like those presented here, is  
820 to identify genomic pathways that are dysregulated leading to obesity pathogenesis, and may result in a  
821 myriad of downstream illnesses. Thus, our findings may enhance the understanding of central obesity  
822 and identify new molecular targets to avert its negative health consequences. Significant cross-trait

823 associations and additional associations observed in the GWAS Catalog are consistent with expected  
824 direction of effect for several traits, i.e. the WHR-increasing allele is associated with higher values of TG,  
825 DBP, fasting insulin, TC, LDL and T2D across many significant variants. However, it is worth noting that  
826 there are some exceptions. For example, rs9469913-A in *UHRF1BP1* is associated with both increased  
827 WHRadjBMI and increased HDL. Also, we identified two variants in *MLXIPL* (rs3812316 and rs35332062),  
828 a well-known lipids-associated locus, in which the WHRadjBMI-increasing allele also increases all lipid  
829 levels, risk for hypertriglyceridemia, SBP and DBP. However, our findings show a significant and negative  
830 association with HbA1C, and nominally significant and negative associations with two-hour glucose,  
831 fasting glucose, and Type 2 diabetes, and potential negative associations with biomarkers for liver  
832 disease (e.g. gamma glutamyl transpeptidase). Other notable exceptions include *ITIH3* (negatively  
833 associated with BMI, HbA1C, LDL and SBP), *DAGLB* (positively associated with HDL), and *STAB1*  
834 (negatively associated with TC, LDL, and SBP in cross-trait associations). Therefore, caution in selecting  
835 pathways for therapeutic targets is warranted; one must look beyond the effects on central adiposity,  
836 but also at the potential cascading effects of related diseases.

837         A seminal finding from this study is the importance of lipid metabolism for body fat distribution.  
838 In fact, pathway analyses that highlight enhanced lipolysis, cross-trait associations with circulating lipid  
839 levels, existing biological evidence from the literature, and knockdown experiments in *Drosophila*  
840 examining triglyceride storage point to novel candidate genes (*ANGPTL4*, *ACVR1C*, *DAGLB*, *MGA*, *RASIP1*,  
841 and *IZUMO1*) and new candidates in known regions (*DNAH10*<sup>10</sup> and *MLXIPL*<sup>14</sup>) related to lipid biology  
842 and its role in fat storage. Newly implicated genes of interest include *ACVR1C*, *MLXIPL*, and *ANGPTL4*, all  
843 of which are involved in lipid homeostasis; all are excellent candidate genes for central adiposity.  
844 Carriers of inactivating mutations in *ANGPTL4* (*Angiopoietin Like 4*), for example, display low triglyceride  
845 levels and low risk of coronary artery disease<sup>63</sup>. *ACVR1C* encodes the activin receptor-like kinase 7  
846 protein (ALK7), a receptor for the transcription factor TGF $\beta$ -1, well known for its central role in growth

847 and development in general<sup>64-68</sup>, and adipocyte development in particular<sup>68</sup>. *ACVR1C* exhibits the highest  
848 expression in adipose tissue, but is also highly expressed in the brain<sup>69-71</sup>. In mice, decreased activity of  
849 *ACVR1C* upregulates PPAR $\gamma$  and C/EBP $\alpha$  pathways and increases lipolysis in adipocytes, thus decreasing  
850 weight and diabetes in mice<sup>69,72,73</sup>. Such activity is suggestive of a role for ALK7 in adipose tissue  
851 signaling and therefore for therapeutic targets for human obesity. *MLXIPL*, also important for lipid  
852 metabolism and postnatal cellular growth, is a transcription factor which activates triglyceride synthesis  
853 genes in a glucose-dependent manner<sup>74,75</sup>. The lead exome variant in this gene is highly conserved, most  
854 likely damaging, and is associated with reduced *MLXIPL* expression in adipose tissue. Furthermore, in a  
855 recent longitudinal, *in vitro* transcriptome analysis of adipogenesis in human adipose-derived stromal  
856 cells, gene expression of *MLXIPL* was up-regulated during the maturation of adipocytes, suggesting a  
857 critical role in the regulation of adipocyte size and accumulation<sup>76</sup>. However, given our observations on  
858 cross-trait associations with variants in *MLXIPL* and diabetes-related traits, development of therapeutic  
859 targets must be approached cautiously.

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

867

868

869 **ACKNOWLEDGEMENTS**

870 A full list of acknowledgements is provided in the **Supplementary Table 17**. This study was completed  
871 as part of the Genetic Investigation of ANthropometric Traits (GIANT) Consortium. This research has been  
872 conducted using the UK Biobank resource. Funding for this project was provided by Aase and Ejner  
873 Danielsens Foundation, Academy of Finland (102318; 123885; 117844; 40758; 211497; 118590; 139635;  
874 129293; 286284; 134309; 126925; 121584; 124282; 129378; 117787; 41071; 137544; 272741), Action  
875 on Hearing Loss (G51), ALK-Abelló A/S (Hørsholm-Denmark), American Heart Association  
876 (13EIA14220013; 13GRNT16490017; 13POST16500011), American Recovery and Reinvestment Act of  
877 2009 (ARRA) Supplement (EY014684-03S1; -04S1), Amgen, André and France Desmarais Montreal Heart  
878 Institute (MHI) Foundation, AstraZeneca, Augustinus Foundation, Australian Government and  
879 Government of Western Australia, Australian Research Council Future Fellowship, Becket Foundation,  
880 Benzon Foundation, Bernard Wolfe Health Neuroscience Endowment, British Heart Foundation  
881 (CH/03/001; RG/14/5/30893; RG/200004; SP/04/002; SP/09/002), BiomarCaRE (278913),  
882 Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research-Germany;  
883 German Center for Diabetes Research (DZD); 01ER1206; 01ER1507; 01ER1206; 01ER1507; FKZ:  
884 01EO1501 (AD2-060E); 01ZZ9603; 01ZZ0103; 01ZZ0403; 03IS2061A; 03Z1CN22; FKZ 01GI1128),  
885 Boehringer Ingelheim Foundation, Boston University School of Medicine, Canada Research Chair  
886 program, Canadian Cancer Society Research Institute, Canadian Institutes of Health Research (MOP-  
887 82893), Cancer Research UK (C864/A14136; A490/A10124; C8197/A16565), Cebu Longitudinal Health  
888 and Nutrition Survey (CLHNS) pilot funds (RR020649; ES010126; DK056350), Center for Non-  
889 Communicable Diseases (Pakistan), Central Society for Clinical Research, Centre National de Génotypage  
890 (Paris-France), CHDI Foundation (Princeton-USA), Chief Scientist Office of the Scottish Government  
891 Health Directorate (CZD/16/6), City of Kuopio and Social Insurance Institution of Finland (4/26/2010),  
892 Clarendon Scholarship, Commission of the European Communities; Directorate C-Public Health

893 (2004310), Copenhagen County, County Council of Dalarna, Curtin University of Technology, Dalarna  
894 University, Danish Centre for Evaluation and Health Technology Assessment, Danish Council for  
895 Independent Research, Danish Diabetes Academy, Danish Heart Foundation, Danish Medical Research  
896 Council-Danish Agency for Science Technology and Innovation, Danish Medical Research Council, Danish  
897 Pharmaceutical Association, Danish Research Council for Independent Research, Dekker scholarship  
898 (2014T001), Dentistry and Health Sciences, Department of Internal Medicine at the University of  
899 Michigan, Diabetes Care System West-Friesland, Diabetes Heart Study (R01 HL6734; R01 HL092301; R01  
900 NS058700), Doris Duke Charitable Foundation Clinical Scientist Development Award (2014105), Doris  
901 Duke Medical Foundation, Dr. Robert Pflieger Stiftung, Dutch Cancer Society (NKI2009-4363), Dutch  
902 Government (NWO 184.021.00; NWO/MaGW VIDI-016-065-318; NWO VICI 453-14-0057; NWO  
903 184.021.007), Dutch Science Organization (ZonMW-VENI Grant 916.14.023), Edith Cowan University,  
904 Education and Sports Research Grant (216-1080315-0302); Croatian Science Foundation (grant 8875),  
905 Else Kröner-Frsenius-Stiftung (2012\_A147), Emil Aaltonen Foundation, Erasmus Medical Center, Erasmus  
906 University (Rotterdam), European Research Council Advanced Principal Investigator Award, European  
907 Research Council (310644; 268834; 323195; SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC; 293574),  
908 Estonian Research Council (IUT20-60), European Union Framework Programme 6  
909 (LSHM\_CT\_2006\_037197; Bloodomics Integrated Project; LSHM-CT-2004-005272; LSHG-CT-2006-  
910 018947), European Union Framework Programme 7 (HEALTH-F2-2013-601456; HEALTH-F2-2012-  
911 279233; 279153; HEALTH-F3-2010-242244; EpiMigrant; 279143; 313010; 305280; HZ2020 633589;  
912 313010; HEALTH-F2-2011-278913; HEALTH-F4-2007- 201413), European Commission (DG XII), European  
913 Community (SOC 98200769 05 F02), European Regional Development Fund to the Centre of Excellence  
914 in Genomics and Translational Medicine (GenTransMed), European Union (QLG1-CT-2001-01252; SOC  
915 95201408 05 F02), EVO funding of the Kuopio University Hospital from Ministry of Health and Social  
916 Affairs (5254), Eye Birth Defects Foundation Inc., Federal Ministry of Science-Germany (01 EA 9401),

917 Finland's Slottery Machine Association, Finnish Academy (255935; 269517), Finnish Cardiovascular  
918 Research Foundation, Finnish Cultural Foundation, Finnish Diabetes Association, Finnish Diabetes  
919 Research Foundation, Finnish Foundation for Cardiovascular Research, Finnish Funding Agency for  
920 Technology and Innovation (40058/07), Finnish Heart Association, Finnish National Public Health  
921 Institute, Fondation Leducq (14CVD01), Food Standards Agency (UK), Framingham Heart Study of the  
922 National Heart Lung and Blood Institute of the National Institutes of Health (HHSN268201500001; N02-  
923 HL-6-4278), FUSION Study (DK093757; DK072193; DK062370; ZIA-HG000024), General Clinical Research  
924 Centre of the Wake Forest School of Medicine (M01 RR07122; F32 HL085989), Genetic Laboratory of the  
925 Department of Internal Medicine-Erasmus MC (the Netherlands Genomics Initiative), Genetics and  
926 Epidemiology of Colorectal Cancer Consortium (NCI CA137088), German Cancer Aid (70-2488-Ha I),  
927 German Diabetes Association, German Research Foundation (CRC 1052 C01; B01; B03), Health and  
928 Retirement Study (R03 AG046398), Health Insurance Foundation (2010 B 131), Health Ministry of  
929 Lombardia Region (Italy), Helmholtz Zentrum München – German Research Center for Environmental  
930 Health, Helse Vest, Home Office (780-TETRA), Hospital Districts of Pirkanmaa; Southern Ostrobothnia;  
931 North Ostrobothnia; Central Finland and Northern Savo, Ib Henriksen Foundation, Imperial College  
932 Biomedical Research Centre, Imperial College Healthcare NHS Trust, Institute of Cancer Research and  
933 The Everyman Campaign, Interuniversity Cardiology Institute of the Netherlands (09.001), Intramural  
934 Research Program of the National Institute on Aging, Italian Ministry of Health (GR-2011-02349604),  
935 Johns Hopkins University School of Medicine (HHSN268200900041C), Juho Vainio Foundation, Kaiser  
936 Foundation Research Institute (HHSN268201300029C), KfH Stiftung Präventivmedizin e.V., KG Jebsen  
937 Foundation, Knut and Alice Wallenberg Foundation (Wallenberg Academy Fellow), Knut och Alice  
938 Wallenberg Foundation (2013.0126), Kuopio Tampere and Turku University Hospital Medical Funds  
939 (X51001), Kuopio University Hospital, Leenaards Foundation, Leiden University Medical Center, Li Ka  
940 Shing Foundation (CML), Ludwig-Maximilians-Universität, Lund University, Lundbeck Foundation,



941 Lundbeckfonden, Marianne and Marcus Wallenberg Foundation, Max Planck Society, Medical Research  
942 Council-UK (G0601966; G0700931; G0000934; MR/L01632X/1; MC\_UU\_12015/1; MC\_PC\_13048;  
943 G9521010D; G1000143; MC\_UU\_12013/1-9; MC\_UU\_12015/1; MC\_PC\_13046; MC\_U106179471;  
944 G0800270, MR/L01341X/1), MEKOS Laboratories (Denmark), Merck & Co Inc., MESA Family (R01-HL-  
945 071205; R01-HL-071051; R01-HL-071250; R01-HL-071251; R01-HL-071252; R01-HL-071258; R01-HL-  
946 071259; UL1-RR-025005), Ministry for Health Welfare and Sports (the Netherlands), Ministry of Cultural  
947 Affairs (Germany), Ministry of Education and Culture of Finland (627;2004-2011), Ministry of Education  
948 Culture and Science (the Netherlands), Ministry of Science and Technology (Taiwan) (MOST 104-2314-B-  
949 075A-006 -MY3), Ministry of Social Affairs and Health in Finland, Montreal Heart Institute Foundation,  
950 MRC-PHE Centre for Environment and Health, Multi-Ethnic Study of Atherosclerosis (MESA) (N01-HC-  
951 95159; N01-HC-95160; N01-HC-95161; N01-HC-95162; N01-HC-95163; N01-HC-95164; N01-HC-95165;  
952 N01-HC-95166; N01-HC-95167; N01-HC-95168; N01-HC-95169), Munich Center of Health Sciences (MC-  
953 Health), Municipality of Rotterdam (the Netherlands) Murdoch University, National Basic Research  
954 Program of China (973 Program 2012CB524900), National Cancer Institute (CA047988; UM1CA182913),  
955 National Cancer Research Institute UK, National Cancer Research Network UK, National Center for  
956 Advancing Translational Sciences (UL1TR001881), National Center for Research Resources (UL1-TR-  
957 000040 and UL1-RR-025005), National Eye Institute of the National Institutes of Health (EY014684, EY-  
958 017337), National Health and Medical Research Council of Australia (403981; 1021105; 572613),  
959 National Heart Lung and Blood Institute (HHSN268201100010C; HHSN268201100011C; and  
960 HHSN268201100012C; HHSN268201100005; HHSN268201100006C; HHSN268201100007C;  
961 HHSN268201100008C; HHSN268201100009; HHSN268201100037C; HHSN268201300046C;  
962 HHSN268201300047C; HHSN268201300048C; HHSN268201300049C; HHSN268201300050C; HL043851;  
963 HL080467; HL094535; HL109946; HHSN268201300025C; HHSN268201300026C; HL119443; HL054464;  
964 HL054457; HL054481; HL087660; HL086694; HL060944; HL061019; HL060919; HL060944; HL061019;

965 N02-HL-6-4278; R21 HL121422-02; R21 HL121422-02; R01 DK089256-05), National Human Genome  
966 Research Institute (HG007112), National Institute for Health Research BioResource Clinical Research  
967 Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's  
968 College London, National Institute for Health Research Comprehensive Biomedical Research Centre  
969 Imperial College Healthcare NHS Trust, National Institute for Health Research (NIHR) (RP-PG-0407-  
970 10371), National Institute of Diabetes and Digestive and Kidney Disease (DK063491; DK097524;  
971 DK085175; DK087914; 1R01DK8925601; 1R01DK106236-01A1), National Institute of Health Research  
972 Senior Investigator, National Institute on Aging (NIA U01AG009740; RC2 AG036495; RC4 AG039029),  
973 National Institute on Minority Health and Health Disparities, National Institutes of Health (NIH)  
974 (1R01HG008983-01; 1R21DA040177-01; 1R01HL092577; R01HL128914; K24HL105780; K01HL116770;  
975 U01 HL072515-06; U01 HL84756; U01HL105198; U01 GM074518; R01 DK089256-05; R01DK075787; R25  
976 CA94880; P30 CA008748; DK078150; TW005596; HL085144; TW008288; R01-HL093029; U01-  
977 HG004729; R01-DK089256; 1R01DK101855-01; 1K99HL130580; T32-GM067553; U01-DK105561; R01-  
978 HL-117078; R01-DK-089256; U01HG008657; U01HG06375; U01AG006781; DK064265; R01DK106621-  
979 01; K23HL114724; NS33335; HL57818; R01-DK089256; 2R01HD057194; U01HG007416; R01DK101855,  
980 R01DK075787, T32 GM096911-05; K01 DK107836; R01DK075787; U01 AG 06781; U01-HG005152,  
981 1F31HG009850-01), National Key R&D Plan of China (2016YFC1304903), Key Project of the Chinese  
982 Academy of Sciences (ZDBS-SSW-DQC-02, ZDRW-ZS-2016-8-1, KJZD-EW-L14-2-2), National Natural  
983 Science Foundation of China (81471013; 30930081; 81170734; 81321062; 81471013), National NIHR  
984 Bioresource, National Science Council (Taiwan) (NSC 102-2314-B-075A-002), Netherlands CardioVascular  
985 Research Initiative (CVON2011-19), Netherlands Heart Foundation, Netherlands Organisation for Health  
986 Research and Development (ZonMW) (113102006), Netherlands Organisation for Scientific Research  
987 (NWO)-sponsored Netherlands Consortium for Healthy Aging (050-060-810), Netherlands Organization  
988 for Scientific Research (184021007), NHMRC Practitioner Fellowship (APP1103329), NIH through the

989 American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419), NIHR Biomedical Research  
990 Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, NIHR  
991 Cambridge Biomedical Research Centre, NIHR Cambridge Biomedical Research Centre, NIHR Health  
992 Protection Research Unit on Health Impact of Environmental Hazards (HPRU-2012-10141), NIHR  
993 Leicester Cardiovascular Biomedical Research Unit, NIHR Official Development Assistance (ODA, award  
994 16/136/68), NIHR Oxford Biomedical Research Centre, the European Union FP7 (EpiMigrant, 279143)  
995 and H2020 programs (iHealth-T2D; 643774), NIHR Senior Investigator, Nordic Centre of Excellence on  
996 Systems Biology in Controlled Dietary Interventions and Cohort Studies (SYSDIET) (070014),  
997 Northwestern University (HHSN268201300027C), Norwegian Diabetes Association, Novartis, Novo  
998 Nordisk Foundation, Nuffield Department of Clinical Medicine Award, Orchid Cancer Appeal, Oxford  
999 Biomedical Research Centre, Paavo Nurmi Foundation, Päivikki and Sakari Sohlberg Foundation, Pawsey  
1000 Supercomputing Centre (funded by Australian Government and Government of Western Australia),  
1001 Peninsula Research Bank-NIHR Exeter Clinical Research Facility, Pfizer, Prostate Cancer Research  
1002 Foundation, Prostate Research Campaign UK (now Prostate Action), Public Health England, QIMR  
1003 Berghofer, Raine Medical Research Foundation, Regione FVG (L.26.2008), Republic of Croatia Ministry of  
1004 Science, Research Centre for Prevention and Health-the Capital Region of Denmark, Research Council of  
1005 Norway, Research Institute for Diseases in the Elderly (RIDE), Research into Ageing, Robert Dawson  
1006 Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston  
1007 Medical Center, Science Live/Science Center NEMO, Scottish Funding Council (HR03006), Sigrid Juselius  
1008 Foundation, Social Insurance Institution of Finland, Singapore Ministry of Health's National Medical  
1009 Research Council (NMRC/STaR/0028/2017), Social Ministry of the Federal State of Mecklenburg-West  
1010 Pomerania, State of Bavaria-Germany, State of Washington Life Sciences Discovery Award (265508) to  
1011 the Northwest Institute of Genetic Medicine, Stroke Association, Swedish Diabetes Foundation (2013-  
1012 024), Swedish Heart-Lung Foundation (20120197; 20120197; 20140422), Swedish Research Council

1013 (2012-1397), Swedish Research Council Strategic Research Network Epidemiology for Health, Swiss  
1014 National Science Foundation (31003A-143914), SystemsX.ch (51RTP0\_151019), Taichung Veterans  
1015 General Hospital (Taiwan) (TCVGH-1047319D; TCVGH-1047311C), Tampere Tuberculosis Foundation,  
1016 TEKES Grants (70103/06; 40058/07), The Telethon Kids Institute, Timber Merchant Vilhelm Bangs  
1017 Foundation, UCL Hospitals NIHR Biomedical Research Centre, UK Department of Health, Université de  
1018 Montréal Beaulieu-Saucier Chair in Pharmacogenomics, University Hospital Regensburg, University of  
1019 Bergen, University of Cambridge, University of Michigan Biological Sciences Scholars Program, University  
1020 of Michigan Internal Medicine Department Division of Gastroenterology, University of Minnesota  
1021 (HHSN268201300028C), University of Notre Dame (Australia), University of Queensland, University of  
1022 Western Australia (UWA), Uppsala Multidisciplinary Center for Advanced Computational Science  
1023 (b2011036), Uppsala University, US Department of Health and Human Services (HHSN268201100046C;  
1024 HHSN268201100001C; HHSN268201100002C; HHSN268201100003C; HHSN268201100004C;  
1025 HHSN271201100004C), UWA Faculty of Medicine, Velux Foundation, Wellcome Trust (083948/B/07/Z;  
1026 084723/Z/08/Z; 090532; 098381; 098497/Z/12/Z; WT098051; 068545/Z/02; WT064890; WT086596;  
1027 WT098017; WT090532; WT098051; WT098017; WT098381; WT098395; 083948; 085475), Western  
1028 Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling  
1029 Facility), Women and Infant's Research Foundation, Yrjö Jahnsson Foundation (56358)

1030

### 1031 **AUTHORSHIP CONTRIBUTIONS**

1032 Writing Group: LAC, RSF, TMF, MG, HMH, JNH, AEJ, TK, ZK, CML, RJFL, YL, KEN, VT, KLY; Data preparation  
1033 group: TA, IBB, TE, SF, MG, HMH, AEJ, TK, DJL, KSL, AEL, RJFL, YL, EM, NGDM, MCMG, PM, MCYN, MAR,  
1034 SS, CS, KS, VT, SV, SMW, TWW, KLY, XZ; WHR meta-analyses: PLA, HMH, AEJ, TK, MG, CML, RJFL, KEN,  
1035 VT, KLY; Pleiotropy working group: GA, MB, JPC, PD, FD, JCF, HMH, SK, HK, HMH, AEJ, CML, DJL, RJFL,  
1036 AM, EM, GM, MIM, PBM, GMP, JRBP, KSR, XS, SW, JW, CJW; Phenome-wide association studies: LB, JCD,

1037 TLE, AG, AM, MIM; Gene-set enrichment analyses: SB, RSF, JNH, ZK, DL, THP; eQTL analyses: CKR, YL,  
1038 KLM; Monogenic and syndromic gene enrichment analyses: HMH, AKM; Fly Obesity Screen: AL, JAP;  
1039 Overseeing of contributing studies: (1958 Birth Cohort) PD; (Airwave) PE; (AMC PAS) GKH; (Amish)  
1040 JRO'C; (ARIC) EB; (ARIC, Add Health) KEN; (BRAVE) EDA, RC; (BRIGHT) PBM; (CARDIA) MF, PJS; (Cebu  
1041 Longitudinal Health and Nutrition Survey) KLM; (CHD Exome + Consortium) ASB, JMMH, DFR, JD; (CHES)  
1042 RV; (Clear/eMERGE (Seattle)) GPJ; (CROATIA\_Korcula) VV, OP, IR; (deCODE) KS, UT; (DHS) DWB;  
1043 (DIACORE) CAB; (DPS) JT, JL, MU; (DRSEXTRA) TAL, RR; (EFSOCH) ATH, TMF; (EGCUT) TE; (eMERGE  
1044 (Seattle)) EBL; (EPIC-Potsdam) MBS, HB; (EpiHealth) EI, PWF; (EXTEND) ATH, TMF; (Family Heart Study)  
1045 IBB; (Fenland, EPIC) RAS; (Fenland, EPIC, InterAct) NJW, CL; (FINRISK) SM; (FINRISK 2007 (T2D) ) PJ, VS;  
1046 (Framingham Heart Study) LAC; (FUSION) MB, FSC; (FVG) PG; (Generation Scotland) CH, BHS; (Genetic  
1047 Epidemiology Network of Arteriopathy (GENOA)) SLRK; (GRAPHIC) NJS; (GSK-STABILITY) DMW, LW,  
1048 HDW; (Health) AL; (HELIC MANOLIS) EZ, GD; (HELIC Pomak) EZ, GD; (HUNT-MI) KH, CJW; (Inter99) TH, TJ;  
1049 (IRASFS) LEW, EKS; (Jackson Heart Study (JHS)) JGW; (KORA S4) KS, IMH; (Leipzig-Adults) MB, PK;  
1050 (LOLIPOP-Exome) JCC, JSK; (LOLIPOP-OmniEE) JCC, JSK; (MESA) JIR, XG; (METSIM) JK, ML; (MONICA-  
1051 Brianza) GC; (Montreal Heart Institute Biobank (MHIBB)) MPD, GL, SdD, JCT; (MORGAM Central  
1052 Laboratory) MP; (MORGAM Data Centre) KK; (OBB) FK; (PCOS) APM, CML; (PIVUS) CML, LL; (PRIME -  
1053 Belfast) FK; (PRIME - Lille) PA; (PRIME - Strasbourg) MM; (PRIME - Toulouse) JF; (PROMIS) DS; (QC) MAR;  
1054 (RISC) BB, EF, MW; (Rotterdam Study I) AGU, MAI; (SEARCH) AMD; (SHIP/SHIP-Trend) MD; (SIBS) DFE;  
1055 (SOLID TIMI-52) DMW; (SORBS) APM, MS, AT; (The Mount Sinai BioMe Biobank) EPB, RJFL; (The NEO  
1056 Study) DOMK; (The NHAPC study, The GBTDS study) XL; (The Western Australian Pregnancy Cohort  
1057 (Raine) Study) CEP, SM; (TwinsUK) TDS; (ULSAM) APM; (Vejle Biobank) IB, CC, OP; (WGHS) DIC, PMR;  
1058 (Women's Health Initiative) PLA; (WTCCC-UKT2D) MIM, KRO; (YFS) TL, OTRa; Genotyping of contributing  
1059 studies: (1958 Birth Cohort) KES; (Airwave) EE, MP SL; (AMC PAS) SS; (Amish) LMYA, JAP; (ARIC) EWD,  
1060 MG; (BBMRI-NL) SHV, LB, CMvD, PIWdB; (BRAVE) EDA; (Cambridge Cancer Studies) JGD; (CARDIA) MF;

1061 (CHD Exome + Consortium) ASB, JMMH, DFR, JD, RY(Clear/eMERGE (Seattle)) GPJ; (CROATIA\_Korcula)  
1062 VV; (DIACORE) CAB, MG; (DPS) AUJ, JL; (DRSEXTRA) PK; (EGCUT) TE; (EPIC-Potsdam) MBS, KM;  
1063 (EpiHealth) EI, PWF; (Family Heart Study) KDT; (Fenland, EPIC) RAS; (Fenland, EPIC, InterAct) NJW, CL;  
1064 (FUSION) NN; (FVG) IG, AM; (Generation Scotland) CH; (Genetic Epidemiology Network of Arteriopathy  
1065 (GENOA)) SLRK, JAS; (GRAPHIC) NJS; (GSK-STABILITY) DMW; (Health) JBJ; (HELIC MANOLIS) LS; (HELIC  
1066 Pomak) LS; (Inter99) TH, NG; (KORA) MMN; (KORA S4) KS, HG; (Leipzig-Adults) AM; (LOLIPOP-Exome)  
1067 JCC, JSK; (LOLIPOP-OmniEE) JCC, JSK; (MESA) JIR, YDIC, KDT; (METSIM) JK, ML; (Montreal Heart Institute  
1068 Biobank (MHIBB)) MPD; (OBB) FK; (PCOS) APM; (PIVUS) CML; (Rotterdam Study I) AGU, CMG, FR; (SDC)  
1069 JMJ, HV; (SEARCH) AIMD; (SOLID TIMI-52) DMW; (SORBS) APM; (The Mount Sinai BioMe Biobank) EPB,  
1070 RJFL, YL, CS; (The NEO Study) RLG; (The NHAPC study, The GBTDS study) XL, HL, YH; (The Western  
1071 Australian Pregnancy Cohort (Raine) Study) CEP, SM; (TUDR) ZA; (TwinsUK) APM; (ULSAM) APM; (WGHS)  
1072 DIC, AYC; (Women's Health Initiative) APR; (WTCCC-UKT2D) MIM; (YFS) TL, LPL; Phenotyping of  
1073 contributing studies: (Airwave) EE; (AMC PAS) SS; (Amish) LM YA; (ARIC) EWD; (ARIC, Add Health) KEN;  
1074 (BBMRI-NL) SHV; (BRAVE) EDA; (BRIGHT) MJC; (CARL) AR, GG; (Cebu Longitudinal Health and Nutrition  
1075 Survey) NRL; (CHES) RV, MT; (Clear/eMERGE (Seattle)) GPJ, AAB; (CROATIA\_Korcula) OP, IR; (DIACORE)  
1076 CAB, BKK; (DPS) AUJ, JL; (EFSOCH) ATH; (EGCUT) EM; (EPIC-Potsdam) HB; (EpiHealth) EI; (EXTEND) ATH;  
1077 (Family Heart Study) MFF; (Fenland, EPIC, InterAct) NJW; (FIN-D2D 2007) LM, MV; (FINRISK) SM;  
1078 (FINRISK 2007 (T2D)) PJ, HS; (Framingham Heart Study) CSF; (Generation Scotland) CH, BHS; (Genetic  
1079 Epidemiology Network of Arteriopathy (GENOA)) SLRK, JAS; (GRAPHIC) NJS; (GSK-STABILITY) LW, HDW;  
1080 (Health) AL, BHT; (HELIC MANOLIS) LS, AEF, ET; (HELIC Pomak) LS, AEF, MK; (HUNT-MI) KH, OH; (Inter99)  
1081 TJ, NG; (IRASFS) LEW, BK; (KORA) MMN; (LASA (BBMRI-NL)) KMAS; (Leipzig-Adults) MB, PK; (LOLIPOP-  
1082 Exome) JCC, JSK; (LOLIPOP-OmniEE) JCC, JSK; (MESA) MA; (Montreal Heart Institute Biobank (MHIBB))  
1083 GL, KSL, VT; (MORGAM Data Centre) KK; (OBB) FK, MN; (PCOS) CML; (PIVUS) LL; (PRIME - Belfast) FK;  
1084 (PRIME - Lille) PA; (PRIME - Strasbourg) MM; (PRIME - Toulouse) JF; (RISC) BB, EF; (Rotterdam Study I)

1085 MAI, CMGFR, MCZ; (SHIP/SHIP-Trend) NF; (SORBS) MS, AT; (The Mount Sinai BioMe Biobank) EPB, YL,  
1086 CS; (The NEO Study) RdM; (The NHAPC study, The GBTDS study) XL, HL, LS, FW; (The Western Australian  
1087 Pregnancy Cohort (Raine) Study) CEP; (TUDR) YJH, WJL; (TwinsUK) TDS, KSS; (ULSAM) VG; (WGHS) DIC,  
1088 PMR; (Women's Health Initiative) APR; (WTCCC-UKT2D) MIM, KRO; (YFS) TL, OTR; Data analysis of  
1089 contributing studies: (1958 Birth Cohort) KES, IN; (Airwave) EE, MPSL; (AMC PAS) SS; (Amish) JRO'C,  
1090 LMYA, JAP; (ARIC, Add Health) KEN, KLY, MG; (BBMRI-NL) LB; (BRAVE) RC, DSA; (BRIGHT) HRW;  
1091 (Cambridge Cancer Studies) JGD, AP, DJT; (CARDIA) MF, LAL; (CARL) AR, DV; (Cebu Longitudinal Health  
1092 and Nutrition Survey) YW; (CHD Exome + Consortium) ASB, JMMH, DFR, RY, PS; (CHES) YJ;  
1093 (CROATIA\_Korcula) VV; (deCODE) VS, GT; (DHS) AJC, PM, MCYN; (DIACORE) CAB, MG; (EFSOCH) HY;  
1094 (EGCUT) TE, RM; (eMERGE (Seattle)) DSC; (ENDO) TK; (EPIC) JHZ; (EPIC-Potsdam) KM; (EpiHealth) SG;  
1095 (EXTEND) HY; (Family Heart Study) MFF; (Fenland) JaL; (Fenland, EPIC) RAS; (Fenland, InterAct) SMW;  
1096 (Finrisk Extremes and QC) SV; (Framingham Heart Study) CTL, NLHC; (FVG) IG; (Generation Scotland) CH,  
1097 JM; (Genetic Epidemiology Network of Arteriopathy (GENOA)) LFB; (GIANT-Analyst) AEJ; (GRAPHIC) NJS,  
1098 NGDM, CPN; (GSK-STABILITY) DMW, AS; (Health) JBJ; (HELIC MANOLIS) LS; (HELIC Pomak) LS; (HUNT-MI)  
1099 WZ; (Inter99) NG; (IRASFS) BK; (Jackson Heart Study (JHS)) LAL, JL; (KORA S4) TWW; (LASA (BBMRI-NL))  
1100 KMAS; (Leipzig-Adults) AM; (LOLIPOP-Exome) JCC, JSK, WZ; (LOLIPOP-OmniEE) JCC, JSK, WZ; (MESA) JIR,  
1101 XG, JY; (METSIM) XS; (Montreal Heart Institute Biobank (MHIBB)) JCT, GL, KSL, VT; (OBB) AM; (PCOS)  
1102 APM, TK; (PIVUS) NR; (PROMIS) AR, WZ; (QC GoT2D/T2D-GENES (FUSION, METSIM, etc)) AEL; (RISC) HY;  
1103 (Rotterdam Study I) CMG, FR; (SHIP/SHIP-Trend) AT; (SOLID TIMI-52) DMW, AS; (SORBS) APM; (The  
1104 Mount Sinai BioMe Biobank) YL, CS; (The NEO Study) RLG; (The NHAPC study, The GBTDS study) XL, HL,  
1105 YH; (The Western Australian Pregnancy Cohort (Raine) Study) CAW; (UK Biobank) ARW; (ULSAM) APM,  
1106 AM; (WGHS) DIC, AYC; (Women's Health Initiative) PLA, JH; (WTCCC-UKT2D) WG; (YFS) LPL.

1107

## 1108 **METHODS**

### 1109 **Studies**

1110           Stage 1 consisted of 74 studies (12 case/control studies, 59 population-based studies, and five  
1111 family studies) comprising 344,369 adult individuals of the following ancestries: 1) European descent (N=  
1112 288,492), 2) African (N= 15,687), 3) South Asian (N= 29,315), 4) East Asian (N=6,800), and 5) Hispanic  
1113 (N=4,075). Stage 1 meta-analyses were carried out in each ancestry separately and in the all ancestries  
1114 group, for both sex-combined and sex-specific analyses. Follow-up analyses were undertaken in 132,177  
1115 individuals of European ancestry from the deCODE anthropometric study and UK Biobank  
1116 (**Supplementary Tables 1-3**). Conditional analyses were performed in the all ancestries and European  
1117 descent groups. Informed consent was obtained for participants by the parent study and protocols  
1118 approved by each study's institutional review boards.

### 1119 **Phenotypes**

1120           For each study, WHR (waist circumference divided by hip circumference) was corrected for age,  
1121 BMI, and the genomic principal components (derived from GWAS data, the variants with MAF >1% on  
1122 the ExomeChip, and ancestry informative markers available on the ExomeChip), as well as any additional  
1123 study-specific covariates (e.g. recruiting center), in a linear regression model. For studies with non-  
1124 related individuals, residuals were calculated separately by sex, whereas for family-based studies sex  
1125 was included as a covariate in models with both men and women. Additionally, residuals for  
1126 case/control studies were calculated separately. Finally, residuals were inverse normal transformed and  
1127 used as the outcome in association analyses. Phenotype descriptives by study are shown in  
1128 **Supplementary Table 3**.

### 1129 **Genotypes and QC**



1130           The majority of studies followed a standardized protocol and performed genotype calling using  
1131 the algorithms indicated in **Supplementary Table 2**, which typically included zCall<sup>3</sup>. For 10 studies  
1132 participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)  
1133 Consortium, the raw intensity data for the samples from seven genotyping centers were assembled into  
1134 a single project for joint calling<sup>4</sup>. Study-specific quality control (QC) measures of the genotyped variants  
1135 were implemented before association analysis (**Supplementary Tables 1-2**). Furthermore, to assess the  
1136 possibility that any significant associations with rare and low-frequency variants could be due to allele  
1137 calling in the smaller studies, we performed a sensitivity meta-analysis including all large studies (>5,000  
1138 participants) and compared to all studies. We found very high concordance for effect sizes, suggesting  
1139 that smaller studies do not bias our results (**Supplementary Fig. 24**).

#### 1140 **Study-level statistical analyses**

1141           Individual cohorts were analyzed for each ancestry separately, in sex-combined and sex-specific  
1142 groups, with either RAREMETALWORKER (<http://genome.sph.umich.edu/wiki/RAREMETALWORKER>) or  
1143 RVTESTs (<http://zhanxw.github.io/rvtests/>), to associate inverse normal transformed WHRadjBMI with  
1144 genotype accounting for cryptic relatedness (kinship matrix) in a linear mixed model. These software  
1145 programs are designed to perform score-statistic based rare-variant association analysis, can  
1146 accommodate both unrelated and related individuals, and provide single-variant results and variance-  
1147 covariance matrices. The covariance matrix captures linkage disequilibrium (LD) relationships between  
1148 markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses<sup>77,78</sup>. Single-  
1149 variant analyses were performed for both additive and recessive models.

#### 1150 **Centralized quality-control**

1151           Individual cohorts identified ancestry population outliers based on 1000 Genome Project phase  
1152 1 ancestry reference populations. A centralized quality-control procedure implemented in EasyQC<sup>79</sup> was

1153 applied to individual cohort association summary statistics to identify cohort-specific problems: (1)  
1154 assessment of possible errors in phenotype residual transformation; (2) comparison of allele frequency  
1155 alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues,  
1156 and (3) examination of quantile-quantile (QQ) plots per study to identify any inflation arising from  
1157 population stratification, cryptic relatedness and genotype biases.

## 1158 **Meta-analyses**

1159 Meta-analyses were carried out in parallel by two different analysts at two sites using  
1160 RAREMETAL<sup>77</sup>. During the meta-analyses, we excluded variants if they had call rate <95%, Hardy-  
1161 Weinberg equilibrium P-value <1x10<sup>-7</sup>, or large allele frequency deviations from reference populations  
1162 (>0.6 for all ancestries analyses and >0.3 for ancestry-specific population analyses). We also excluded  
1163 from downstream analyses markers not present on the Illumina ExomeChip array 1.0, variants on the Y-  
1164 chromosome or the mitochondrial genome, indels, multiallelic variants, and problematic variants based  
1165 on the Blat-based sequence alignment analyses. Significance for single-variant analyses was defined at  
1166 an array-wide level ( $P < 2 \times 10^{-7}$ ). For all suggestive significant variants from Stage 1, we tested for  
1167 significant sex differences. We calculated  $P_{\text{sexhet}}$  for each SNP, testing for difference between women-  
1168 specific and men-specific beta estimates and standard errors using EasyStrata<sup>11,80</sup>. Each SNP that  
1169 reached  $P_{\text{sexhet}} < 0.05/\#$  of variants tested (70 variants brought forward from Stage 1,  $P_{\text{sexhet}} < 7.14 \times 10^{-4}$ )  
1170 was considered significant. Additionally, while each individual study was asked to perform association  
1171 analyses stratified by race/ethnicity, and adjust for population stratification, all study-specific summary  
1172 statistics were meta-analyzed together for our all ancestry meta-analyses. To investigate potential  
1173 heterogeneity across ancestries, we did examine ancestry-specific meta-analysis results for our top 70  
1174 variants from stage 1, and found no evidence of significant across-ancestry heterogeneity observed for  
1175 any of our top variants ( $I^2$  values noted in **Supplementary Data 1-3**).

1176 For the gene-based analyses, we applied two sets of criteria to select variants with a MAF<5%  
1177 within each ancestry based on coding variant annotation from five prediction algorithms (PolyPhen2,  
1178 HumDiv and HumVar, LRT, MutationTaster, and SIFT)<sup>80,81</sup>. Our broad gene-based tests included  
1179 nonsense, stop-loss, splice site, and missense variants annotated as damaging by at least one algorithm  
1180 mentioned above. Our strict gene-based tests included only nonsense, stop-loss, splice site, and  
1181 missense variants annotated as damaging by all five algorithms. These analyses were performed using  
1182 the sequence kernel association test (SKAT) and variable threshold (VT) methods. Statistical significance  
1183 for gene-based tests was set at a Bonferroni-corrected threshold of  $P < 2.5 \times 10^{-6}$  (0.05/~20,000 genes). All  
1184 gene-based tests were performed in RAREMETAL<sup>77</sup>.

### 1185 **Genomic inflation**

1186 We observed a marked genomic inflation of the test statistics even after controlling for  
1187 population stratification (linear mixed model) arising mainly from common markers;  $\lambda_{GC}$  in the primary  
1188 meta-analysis (combined ancestries and combined sexes) was 1.06 and 1.37 for all and only common  
1189 coding and splice site markers considered herein, respectively (**Supplementary Figures 3, 7 and 13,**  
1190 **Supplementary Table 16**). Such inflation is expected for a highly polygenic trait like WHRadjBMI, for  
1191 studies using a non-random set of variants across the genome, and is consistent with our very large  
1192 sample size<sup>79,82,83</sup>.

### 1193 **Conditional analyses**

1194 The RAREMETAL R-package<sup>77</sup> was used to identify independent WHRadjBMI association signals  
1195 across all ancestries and European meta-analysis results. RAREMETAL performs conditional analyses by  
1196 using covariance matrices to distinguish true signals from the shadows of adjacent significant variants in  
1197 LD. First, we identified the lead variants ( $P < 2 \times 10^{-7}$ ) based on a 1Mb window centered on the most  
1198 significantly associated variant. We then conditioned on the lead variants in RAREMETAL and kept new

1199 lead signals at  $P < 2 \times 10^{-7}$  for conditioning in a second round of analysis. The process was repeated until no  
1200 additional signal emerged below the pre-specified P-value threshold ( $P < 2 \times 10^{-7}$ ).

1201 To test if the associations detected were independent of the previously published WHRadjBMI  
1202 variants<sup>10,14,16</sup>, we performed conditional analyses in the stage 1 discovery set if the GWAS variant or its  
1203 proxy ( $r^2 \geq 0.8$ ) was present on the ExomeChip using RAREMETAL<sup>77</sup>. All variants identified in our meta-  
1204 analysis and the previously published variants were also present in the UK Biobank dataset<sup>84</sup>. This  
1205 dataset was used as a replacement dataset if a good proxy was not present on the ExomeChip as well as  
1206 a replication dataset for the variants present on the ExomeChip. All conditional analyses in the UK  
1207 Biobank dataset were performed using SNPTTEST<sup>85-87</sup>. The conditional analyses were carried out  
1208 reciprocally, conditioning on the ExomeChip variant and then the previously published variant. An  
1209 association was considered independent of the previously published association if there was a  
1210 statistically significant association detected prior to the conditional analysis ( $P < 2 \times 10^{-7}$ ) with both the  
1211 exome chip variant and the previously published variant, and the observed association with both or  
1212 either of the variants disappeared upon conditional analysis ( $P > 0.05$ ). A conditional p-value between  
1213  $9 \times 10^{-6}$  and 0.05 was considered inconclusive. However, a conditional p-value  $< 9 \times 10^{-6}$  was also  
1214 considered suggestive.

1215

## 1216 **Stage 2 meta-analyses**

1217 In our Stage 2, we sought to validate a total of 70 variants from Stage 1 that met  $P < 2 \times 10^{-6}$  in two  
1218 independent studies, the UK Biobank (Release 1<sup>84</sup>) and Iceland (deCODE), comprising 119,572 and  
1219 12,605 individuals, respectively (Supplementary Tables 1-3). The same QC and analytical methodology  
1220 were used for these studies. Genotyping, study descriptions and phenotype descriptives are provided in  
1221 **Supplementary Tables 1-3**. For the combined analysis of Stage 1 plus 2, we used the inverse-variance  
1222 weighted fixed effects meta-analysis method. Significant associations were defined as those nominally

1223 significant ( $P < 0.05$ ) in the Stage 2 study and for the combined meta-analysis (Stage 1 plus Stage 2)

1224 significance was set at  $P < 2 \times 10^{-7}$  ( $0.05 / \sim 250,000$  variants).

### 1225 **Pathway enrichment analyses: EC-DEPICT**

1226 We adapted DEPICT, a gene set enrichment analysis method for GWAS data, for use with the  
1227 ExomeChip ('EC-DEPICT'); this method is also described in a companion manuscript<sup>22</sup>. DEPICT's primary  
1228 innovation is the use of "reconstituted" gene sets, where many different types of gene sets (e.g.  
1229 canonical pathways, protein-protein interaction networks, and mouse phenotypes) were extended  
1230 through the use of large-scale microarray data (see Pers et al.<sup>21</sup> for details). EC-DEPICT computes p-  
1231 values based on Swedish ExomeChip data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania  
1232 (ANDIS), and Scania Diabetes Registry (SDR) cohorts,  $N = 11,899$ ) and, unlike DEPICT, takes as input only  
1233 the genes directly containing the significant (coding) variants rather than all genes within a specified  
1234 amount of linkage disequilibrium (see **Supplementary Note 2**).

1235 Two analyses were performed for WHRadjBMI ExomeChip: one with all variants  $p < 5 \times 10^{-4}$  (49  
1236 significant gene sets in 25 meta-gene sets,  $FDR < 0.05$ ) and one with all variants  $> 1$  Mb from known  
1237 GWAS loci<sup>10</sup> (26 significant gene sets in 13 meta-gene sets,  $FDR < 0.05$ ). Affinity propagation clustering<sup>88</sup>  
1238 was used to group highly correlated gene sets into "meta-gene sets"; for each meta-gene set, the  
1239 member gene set with the best p-value was used as representative for purposes of visualization (see  
1240 Supplementary Note). DEPICT for ExomeChip was written using the Python programming language, and  
1241 the code can be found at <https://github.com/RebeccaFine/obesity-ec-depict>.

### 1242 **Pathway enrichment analyses: PASCAL**

1243 We also applied the PASCAL pathway analysis tool<sup>23</sup> to exome-wide association summary  
1244 statistics from Stage 1 for all coding variants. The method derives gene-based scores (both SUM and  
1245 MAX statistics) and subsequently tests for over-representation of high gene scores in predefined

1246 biological pathways. We used standard pathway libraries from KEGG, REACTOME and BIOCARTA, and  
1247 also added dichotomized ( $Z$ -score $>3$ ) reconstituted gene sets from DEPICT<sup>21</sup>. To accurately estimate  
1248 SNP-by-SNP correlations even for rare variants, we used the UK10K data (TwinsUK<sup>89</sup> and ALSPAC<sup>90</sup>  
1249 studies,  $N=3781$ ). In order to separate the contribution of regulatory variants from the coding variants,  
1250 we also applied PASCAL to association summary statistics of only regulatory variants (20 kb upstream)  
1251 and regulatory+coding variants from the Shungin et al<sup>10</sup> study. In this way, we could comment on what is  
1252 gained by analyzing coding variants available on ExomeChip arrays. We performed both MAX and SUM  
1253 estimations for pathway enrichment. MAX is more sensitive to genesets driven primarily by a single  
1254 signal, while SUM is better when there are multiple variant associations in the same gene.

### 1255 **Monogenic obesity enrichment analyses**

1256 We compiled two lists consisting of 31 genes with strong evidence that disruption causes  
1257 monogenic forms of insulin resistance or diabetes; and 8 genes with evidence that disruption causes  
1258 monogenic forms of lipodystrophy. To test for enrichment of association, we conducted simulations by  
1259 matching each gene with others based on gene length and number of variants tested, to create a  
1260 matched set of genes. We generated 1,000 matched gene sets from our data, and assessed how often  
1261 the number of variants exceeding set significance thresholds was greater than in our monogenic obesity  
1262 gene set.

### 1263 **Variance explained**

1264 We estimated the phenotypic variance explained by the association signals in Stage 1 all  
1265 ancestries analyses for men, women, and combined sexes<sup>91</sup>. For each associated region, we pruned  
1266 subsets of SNPs within 500 kb, as this threshold was comparable with previous studies, of the SNPs with  
1267 the lowest P-value and used varying P value thresholds (ranging from  $2 \times 10^{-7}$  to 0.02) from the combined  
1268 sexes results. Additionally, we examined all variants and independent variants across a range of MAF

1269 thresholds. The variance explained by each subset of SNPs in each strata was estimated by summing the  
1270 variance explained by the individual top coding variants. For the comparison of variance explained  
1271 between men and women, we tested for the significance of the differences assuming that the weighted  
1272 sum of chi-squared distributed variables tend to a Gaussian distribution ensured by Lyapunov's central  
1273 limit theorem.<sup>91,92</sup>

#### 1274 **Cross-trait lookups**

1275 To carefully explore the relationship between WHRadjBMI and related cardiometabolic,  
1276 anthropometric, and reproductive traits, association results for the 51 WHRadjBMI coding SNPs were  
1277 requested from existing or on-going meta-analyses from 7 consortia, including ExomeChip data from  
1278 GIANT (BMI, height), Global Lipids Genetics Consortium Results (GLGC) (total cholesterol, triglycerides,  
1279 HDL-cholesterol, LDL-cholesterol), International Consortium for Blood Pressure (IBPC)<sup>93</sup> (systolic and  
1280 diastolic blood pressure), Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC)  
1281 (glycemic traits), and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (type 2  
1282 diabetes).<sup>22,25-29</sup> For coronary artery disease, we accessed 1000 Genomes Project-imputed GWAS data  
1283 released by CARDIoGRAMplusC4D<sup>94</sup> and for the ReproGen consortium (age at menarche and  
1284 menopause) we used a combination of ExomeChip and 1000 Genomes Project-Imputed GWAS data.  
1285 Heatmaps were generated in R v3.3.2 using gplots (<https://CRAN.R-project.org/package=gplots>). We  
1286 used Euclidean distance based on p-value and direction of effect and complete linkage clustering for the  
1287 dendrograms.

#### 1288 **GWAS Catalog Lookups**

1289 In order to determine if significant coding variants were associated with any related  
1290 cardiometabolic and anthropometric traits, we also searched the NHGRI-EBI GWAS Catalog for previous  
1291 variant-trait associations near our lead SNPs (+/- 500 kb). We used PLINK to calculate LD for variants

1292 using ARIC study European participants. All SNVs within the specified regions with an  $r^2$  value  $> 0.7$  were  
1293 retained from NHGRI-EBI GWAS Catalog for further evaluation<sup>37</sup>. Consistent direction of effect was  
1294 based on WHR-increasing allele, LD, and allele frequency. Therefore, when a GWAS Catalog variant was  
1295 not identical or in high LD ( $r^2 > 0.9$ ) with the WHR variant, and MAF  $> 0.45$ , we do not comment on  
1296 direction of effect.

### 1297 **Body-fat percentage associations**

1298 We performed body fat percent and truncal fat percent look-up of 48 of the 56 identified  
1299 variants (tables 1 and 2) that were available in the UK Biobank, Release 1<sup>84</sup>, data (notably some of the  
1300 rare variants in table 1 and 2 were not available) to further characterize their effects on WHRadjBMI.  
1301 Genome-wide association analyses for body fat percent and truncal fat percent were carried out in the  
1302 UK Biobank. Prior to analysis, phenotype data were filtered to exclude pregnant or possibly pregnant  
1303 women, individuals with body mass index  $< 15$ , and without genetically confirmed European ancestry,  
1304 resulting in a sample size of 120,286. Estimated measures of body fat percent and truncal fat percent  
1305 were obtained using the Tanita BC418MA body composition analyzer (Tanita, Tokyo, Japan). Individuals  
1306 were not required to fast and did not follow any specific instructions prior to the bioimpedance  
1307 measurements. SNPTEST was used to perform the analyses based on residuals adjusted for age, 15  
1308 principle components, assessment center and the genotyping chip<sup>85</sup>.

### 1309 **Collider bias**

1310 In order to evaluate SNPs for possible collider bias<sup>18</sup>, we used results from a recent association  
1311 analysis from GIANT on BMI<sup>25</sup>. For each significant SNP identified in our additive models, WHRadjBMI  
1312 associations were corrected for potential bias due to associations between each variant and BMI (See  
1313 **Supplementary Note 1** for additional details). Variants were considered robust against collider bias if



1314 they met Bonferroni-corrected significance following correction ( $P_{\text{corrected}} < 9.09 \times 10^{-4}$ , 0.05/55 variants  
1315 examined).

### 1316 **Drosophila RNAi knockdown experiments**

1317 For each gene in which coding variants were associated with WHRadjBMI in the final combined  
1318 meta-analysis ( $P < 2 \times 10^{-7}$ ), its corresponding Drosophila orthologues were identified in the Ensembl  
1319 ortholog database ([www.ensembl.org](http://www.ensembl.org)), when available. Drosophila triglyceride content values were  
1320 mined from a publicly available genome-wide fat screen data set<sup>45</sup> to identify potential genes for follow-  
1321 up knockdowns. Estimated values represent fractional changes in triglyceride content in adult male flies.  
1322 Data are from male progeny resulting from crosses of male UAS-RNAi flies from the Vienna Drosophila  
1323 Resource Center (VDRC) and Hsp70-GAL4; Tub-GAL8ts virgin females. Two-to-five-day-old males were  
1324 sorted into groups of 20 and subjected to two one-hour wet heatshocks four days apart. On the seventh  
1325 day, flies were picked in groups of eight, manually crushed and sonicated, and the lysates heat-  
1326 inactivated for 10 min in a thermocycler at 95 °C. Centrifuge-cleared supernatants were then used for  
1327 triglyceride (GPO Trinder, Sigma) and protein (Pierce) determination. Triglyceride values from these  
1328 adult-induced ubiquitous RNAi knockdown individuals were normalized to those obtained in parallel  
1329 from non-heatshocked progeny from the very same crosses. The screen comprised one to three  
1330 biological replicates. We followed up each gene with a >0.2 increase or >0.4 decrease in triglyceride  
1331 content.

1332 Orthologues for two genes were brought forward for follow-up, *DNAH10* and *PLXND1*. For both  
1333 genes, we generated adipose tissue (*cg-Gal4*) and neuronal (*elav-Gal4*) specific RNAi-knockdown crosses  
1334 to knockdown transcripts in a tissue specific manner, leveraging upstream activation sequence (UAS)-  
1335 inducible short-hairpin knockdown lines, available through the VDRC (Vienna *Drosophila* Resource  
1336 Center). Specifically, *elav-Gal4*, which drives expression of the RNAi construct in post mitotic neurons  
1337 starting at embryonic stages all the way to adulthood, was used. *Cg* drives expression in the fat body and

1338 hemocytes starting at embryonic stage 12, all the way to adulthood. We crossed male UAS-RNAi flies  
1339 and elav-GAL4 or CG-GAL4 virgin female flies. All fly experiments were carried out at 25°C. Five-to-  
1340 seven-day-old males were sorted into groups of 20, weighed and homogenated in PBS with 0.05%  
1341 Tween with Lysing Matrix D in a beadshaker. The homogenate was heat-inactivated for 10 min in a  
1342 thermocycler at 70°C. 10 $\mu$ l of the homogenate was subsequently used in a triglyceride assay (Sigma,  
1343 Serum Triglyceride Determination Kit) which was carried out in duplicate according to protocol, with one  
1344 alteration: the samples were cleared of residual particulate debris by centrifugation before absorbance  
1345 reading. Resulting triglyceride values were normalized to fly weight and larval/population density. We  
1346 used the non-parametric Kruskal-Wallis test to compare wild type with knockdown lines.

#### 1347 **Expression quantitative trait loci (eQTLs) analysis**

1348 We queried the significant variant (Exome coding SNPs)-gene pairs associated with eGenes  
1349 across five metabolically relevant tissues (skeletal muscle, subcutaneous adipose, visceral adipose, liver  
1350 and pancreas) with at least 70 samples in the GTEx database<sup>46</sup>. For each tissue, variants were selected  
1351 based on the following thresholds: the minor allele was observed in at least 10 samples, and the minor  
1352 allele frequency was  $\geq 0.01$ . eGenes, genes with a significant eQTL, are defined on a false discovery rate  
1353 (FDR)<sup>95</sup> threshold of  $\leq 0.05$  of beta distribution-adjusted empirical p-value from FastQTL. Nominal p-  
1354 values were generated for each variant-gene pair by testing the alternative hypothesis that the slope of  
1355 a linear regression model between genotype and expression deviates from 0. To identify the list of all  
1356 significant variant-gene pairs associated with eGenes, a genome-wide empirical p-value threshold<sup>64</sup>, pt,  
1357 was defined as the empirical p-value of the gene closest to the 0.05 FDR threshold. pt was then used to  
1358 calculate a nominal p-value threshold for each gene based on the beta distribution model (from  
1359 FastQTL) of the minimum p-value distribution  $f(p_{min})$  obtained from the permutations for the gene. For  
1360 each gene, variants with a nominal p-value below the gene-level threshold were considered significant  
1361 and included in the final list of variant-gene pairs<sup>64</sup>. For each eGene, we also listed the most significantly

1362 associated variants (eSNP). Only these exome SNPs with  $r^2 > 0.8$  with eSNPs were considered for the  
1363 biological interpretation (Supplementary eQTL GTEX).

1364 We also performed cis-eQTL analysis in 770 METSIM subcutaneous adipose tissue samples as  
1365 described in Civelek, et al.<sup>96</sup> A false discovery rate (FDR) was calculated using all p-values from the cis-  
1366 eQTL detection in the q-value package in R. Variants associated with nearby genes at an FDR less than  
1367 1% were considered to be significant (equivalent p-value  $< 2.46 \times 10^{-4}$ ).

1368 For loci with more than one microarray probeset of the same gene associated with the exome  
1369 variant, we selected the probeset that provided the strongest LD  $r^2$  between the exome variant and the  
1370 eSNP. In reciprocal conditional analysis, we conditioned on the lead exome variant by including it as a  
1371 covariate in the cis-eQTL detection and reporting the p-value of the eSNP and vice versa. We considered  
1372 the signals to be coincident if both the lead exome variant and the eSNP were no longer significant after  
1373 conditioning on the other and the variants were in high pairwise LD ( $r^2 > 0.80$ ).

1374 For loci that also harbored reported GWAS variants, we performed reciprocal conditional  
1375 analysis between the GWAS lead variant and the lead eSNP. For loci with more than one reported GWAS  
1376 variant, the GWAS lead variant with the strongest LD  $r^2$  with the lead eSNP was reported.

### 1377 **Penetrance analysis**

1378 Phenotype and genotype data from the UK Biobank (UKBB) were used for the penetrance  
1379 analysis. Three of 16 rare and low frequency variants ( $MAF \leq 1\%$ ) detected in the final Stage 1 plus 2  
1380 meta-analysis were available in the UKBB and had relatively larger effect sizes ( $>0.90$ ). The phenotype  
1381 data for these three variants were stratified with respect to waist-to-hip ratio (WHR) using the World  
1382 Health Organization (WHO) guidelines. These guidelines consider women and men with WHR greater  
1383 than 0.85 and 0.90 as obese, respectively. Genotype and allele counts were obtained for the available  
1384 variants and these were used to calculate the number of carriers of the minor allele. The number of

1385 carriers for women, men and all combined was then compared between two strata (obese vs. non-  
1386 obese) using a  $\chi^2$  test. The significance threshold was determined by using a Bonferroni correction for  
1387 the number of tests performed ( $0.05/9=5.5 \times 10^{-3}$ )).

#### 1388 **DATA AVAILABILITY**

1389 Summary statistics of all analyses are available at <https://www.broadinstitute.org/collaboration/giant/>.

1390

1391 **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 ( $\beta$  (SE)= 0.0156 (0.0024), P-value= $9.16 \times 10^{-11}$ ). *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<sup>97-101</sup>. Mutations in this gene are associated with Moebius syndrome<sup>102-105</sup>, and persistent truncus arteriosus<sup>99,106</sup>. *PLXND1* is involved in angiogenesis as part of the SEMA and VEGF signalling pathways<sup>107-110</sup>. *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<sup>101</sup>. *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 ( $\beta$  (SE)= 0.0652 (0.0105), P-value=  $4.81 \times 10^{-10}$ ). *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<sup>64-68</sup>, including adipocytes<sup>68</sup>. Mouse *Acvr1c* decreases secretion of insulin and

is involved in lipid storage<sup>69,72,73,69,72,73,111</sup>. *ACVR1C* exhibits the highest expression in adipose tissue, but is also highly expressed in the brain (GTEx)<sup>69-71</sup>. Expression is associated with body fat, carbohydrate metabolism and lipids in both obese and lean individuals<sup>70</sup>. *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 ( $\beta$  (SE) = 0.258 (0.049), P-value=  $1.38 \times 10^{-07}$ ). 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<sup>112</sup>. *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 ( $\beta$  (SE) = 0.064 (0.011) P-value=  $1.20 \times 10^{-09}$ ). *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<sup>113,114</sup>, and insulin sensitivity<sup>115</sup>. Angptl4-deficient mice have hypotriglyceridemia and

increased lipoprotein lipase (LPL) activity, while transgenic mice overexpressing *Angptl4* in the liver have higher plasma triglyceride levels and decreased LPL activity<sup>116</sup>. The major allele of rs116843064 has been previously associated with increased risk of coronary heart disease and increased TG<sup>63</sup>. *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 ( $\beta$  (SE)=0.017 (0.002), P-value=3.9x10<sup>-15</sup>). This variant is independent of the previously reported GWAS signal in the *RREB1* region (rs1294410; 6:6738752<sup>10</sup>). 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<sup>117-119</sup>. The ras responsive transcription factor *RREB1* is a candidate gene for type 2 diabetes associated end-stage kidney disease<sup>118</sup>. 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 ( $\beta$  (SE)= 0.0136 (0.0025), P-value=6.24x10<sup>-8</sup>). *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)<sup>120</sup>. The *DAGLB* variant, rs702485 (7:6449272, r<sup>2</sup>= 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<sup>121</sup>. *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 ( $\beta$  (SE)= 0.02 (0.0033), P-value= $1.78 \times 10^{-9}$ ;  $\beta$  (SE)= 0.0213 (0.0034), P-value= $1.98 \times 10^{-10}$ ). 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<sup>74,75</sup>. 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<sup>74,122,123</sup>. 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 ( $\beta$  (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<sup>124</sup>, may be important in regulating adipocyte differentiation<sup>125-127</sup>, plays an important role in regulating adiposity and energy balance<sup>128</sup>. *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 ( $\beta$ (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<sup>129</sup>. *TBX15* negatively controls depot-specific adipocyte differentiation and function<sup>130</sup> and regulates glycolytic myofiber identity and muscle metabolism<sup>131</sup>. *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).

## 1392 REFERENCES

- 1393 1. Pischon, T. *et al.* General and abdominal adiposity and risk of death in Europe. *N Engl J Med* **359**,  
1394 2105-20 (2008).

- 1395 2. Wang, Y., Rimm, E.B., Stampfer, M.J., Willett, W.C. & Hu, F.B. Comparison of abdominal  
1396 adiposity and overall obesity in predicting risk of type 2 diabetes among men. *Am J Clin Nutr* **81**,  
1397 555-63 (2005).
- 1398 3. Canoy, D. Distribution of body fat and risk of coronary heart disease in men and women. *Curr*  
1399 *Opin Cardiol* **23**, 591-8 (2008).
- 1400 4. Snijder, M.B. *et al.* Associations of hip and thigh circumferences independent of waist  
1401 circumference with the incidence of type 2 diabetes: the Hoorn Study. *Am J Clin Nutr* **77**, 1192-7  
1402 (2003).
- 1403 5. Yusuf, S. *et al.* Obesity and the risk of myocardial infarction in 27,000 participants from 52  
1404 countries: a case-control study. *Lancet* **366**, 1640-9 (2005).
- 1405 6. Mason, C., Craig, C.L. & Katzmarzyk, P.T. Influence of central and extremity circumferences on  
1406 all-cause mortality in men and women. *Obesity (Silver Spring)* **16**, 2690-5 (2008).
- 1407 7. Karpe, F. & Pinnick, K.E. Biology of upper-body and lower-body adipose tissue--link to whole-  
1408 body phenotypes. *Nat Rev Endocrinol* **11**, 90-100 (2015).
- 1409 8. Manolopoulos, K.N., Karpe, F. & Frayn, K.N. Gluteofemoral body fat as a determinant of  
1410 metabolic health. *Int J Obes (Lond)* **34**, 949-59 (2010).
- 1411 9. Emdin, C.A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2  
1412 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626-634 (2017).
- 1413 10. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution.  
1414 *Nature* **518**, 187-96 (2015).
- 1415 11. Winkler, T.W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size  
1416 and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
- 1417 12. Wen, W. *et al.* Genome-wide association studies in East Asians identify new loci for waist-hip  
1418 ratio and waist circumference. *Sci Rep* **6**, 17958 (2016).

- 1419 13. Gao, C. *et al.* A Comprehensive Analysis of Common and Rare Variants to Identify Adiposity Loci  
1420 in Hispanic Americans: The IRAS Family Study (IRASFS). *PLoS One* **10**, e0134649 (2015).
- 1421 14. Graff, M. *et al.* Genome-wide physical activity interactions in adiposity - A meta-analysis of  
1422 200,452 adults. *PLoS Genet* **13**, e1006528 (2017).
- 1423 15. Justice, A.E. *et al.* Genome-wide meta-analysis of 241,258 adults accounting for smoking  
1424 behaviour identifies novel loci for obesity traits. *Nat Commun* **8**, 14977 (2017).
- 1425 16. Ng, M.C.Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of  
1426 genome-wide association studies in individuals of African ancestry: African Ancestry  
1427 Anthropometry Genetics Consortium. *PLoS Genet* **13**, e1006719 (2017).
- 1428 17. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology.  
1429 *Nature* **518**, 197-206 (2015).
- 1430 18. Aschard, H., Vilhjalmsson, B.J., Joshi, A.D., Price, A.L. & Kraft, P. Adjusting for heritable  
1431 covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet* **96**,  
1432 329-39 (2015).
- 1433 19. Day, F.R., Loh, P.R., Scott, R.A., Ong, K.K. & Perry, J.R. A Robust Example of Collider Bias in a  
1434 Genetic Association Study. *Am J Hum Genet* **98**, 392-3 (2016).
- 1435 20. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-  
1436 analysis for rare variants. *Bioinformatics* **30**, 2828-9 (2014).
- 1437 21. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted  
1438 gene functions. *Nat Commun* **6**, 5890 (2015).
- 1439 22. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**,  
1440 186-190 (2017).

- 1441 23. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous  
1442 Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput*  
1443 *Biol* **12**, e1004714 (2016).
- 1444 24. Kawai, M., de Paula, F.J. & Rosen, C.J. New insights into osteoporosis: the bone-fat connection. *J*  
1445 *Intern Med* **272**, 317-29 (2012).
- 1446 25. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways  
1447 that control energy intake and expenditure in obesity. *Nat Genet* **50**, 26-41 (2018).
- 1448 26. Liu, D.J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. **49**, 1758-  
1449 1766 (2017).
- 1450 27. Kraja, A.T. *et al.* New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000  
1451 Individuals. *Circ Cardiovasc Genet* **10**(2017).
- 1452 28. Mahajan, A. *et al.* Identification and functional characterization of G6PC2 coding variants  
1453 influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet*  
1454 **11**, e1004876 (2015).
- 1455 29. Manning, A. *et al.* A Low-Frequency Inactivating AKT2 Variant Enriched in the Finnish Population  
1456 Is Associated With Fasting Insulin Levels and Type 2 Diabetes Risk. *Diabetes* **66**, 2019-2032  
1457 (2017).
- 1458 30. Zhao, W. *et al.* Identification of new susceptibility loci for type 2 diabetes and shared etiological  
1459 pathways with coronary heart disease. **49**, 1450-1457 (2017).
- 1460 31. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture  
1461 and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
- 1462 32. Ng, M.C. *et al.* Meta-analysis of genome-wide association studies in African Americans provides  
1463 insights into the genetic architecture of type 2 diabetes. *PLoS Genet* **10**, e1004517 (2014).

- 1464 33. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic  
1465 architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-44 (2014).
- 1466 34. Saxena, R. *et al.* Genome-wide association study identifies a novel locus contributing to type 2  
1467 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* **62**, 1746-55 (2013).
- 1468 35. Cook, J.P. & Morris, A.P. Multi-ethnic genome-wide association study identifies novel locus for  
1469 type 2 diabetes susceptibility. *Eur J Hum Genet* **24**, 1175-80 (2016).
- 1470 36. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale  
1471 association analysis. *Nat Genet* **42**, 579-89 (2010).
- 1472 37. Burdett, T. *et al.* The NHGRI-EBI Catalog of published genome-wide association studies. v1.0 edn  
1473 Vol. 2015 (2015).
- 1474 38. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association  
1475 loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
- 1476 39. Lutoslawska, G. *et al.* Relationship between the percentage of body fat and surrogate indices of  
1477 fatness in male and female Polish active and sedentary students. *J Physiol Anthropol* **33**, 10  
1478 (2014).
- 1479 40. Verma, M., Rajput, M., Sahoo, S.S., Kaur, N. & Rohilla, R. Correlation between the percentage of  
1480 body fat and surrogate indices of obesity among adult population in rural block of Haryana. *J*  
1481 *Family Med Prim Care* **5**, 154-9 (2016).
- 1482 41. Pereira, P.F. *et al.* [Measurements of location of body fat distribution: an assessment of  
1483 colinearity with body mass, adiposity and stature in female adolescents]. *Rev Paul Pediatr* **33**,  
1484 63-71 (2015).
- 1485 42. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic  
1486 disease risk. *Nat Commun* **7**, 10495 (2016).

- 1487 43. Chambers, J.C. *et al.* Common genetic variation near MC4R is associated with waist  
1488 circumference and insulin resistance. *Nat Genet* **40**, 716-8 (2008).
- 1489 44. Neale, K.T. *et al.* Contribution of common non-synonymous variants in PCSK1 to body mass index  
1490 variation and risk of obesity: a systematic review and meta-analysis with evidence from up to  
1491 331 175 individuals. *Hum Mol Genet* **24**, 3582-94 (2015).
- 1492 45. Pospisilik, J.A. *et al.* Drosophila genome-wide obesity screen reveals hedgehog as a determinant  
1493 of brown versus white adipose cell fate. *Cell* **140**, 148-60 (2010).
- 1494 46. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:  
1495 multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
- 1496 47. Baraille, F., Planchais, J., Dentin, R., Guilmeau, S. & Postic, C. Integration of ChREBP-Mediated  
1497 Glucose Sensing into Whole Body Metabolism. *Physiology (Bethesda)* **30**, 428-37 (2015).
- 1498 48. Kursawe, R. *et al.* Decreased transcription of ChREBP-alpha/beta isoforms in abdominal  
1499 subcutaneous adipose tissue of obese adolescents with prediabetes or early type 2 diabetes:  
1500 associations with insulin resistance and hyperglycemia. *Diabetes* **62**, 837-44 (2013).
- 1501 49. Lotta, L.A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage  
1502 capacity in the pathogenesis of human insulin resistance. *Nat Genet* **49**, 17-26 (2017).
- 1503 50. Cargill, M. *et al.* A large-scale genetic association study confirms IL12B and leads to the  
1504 identification of IL23R as psoriasis-risk genes. *Am J Hum Genet* **80**, 273-90 (2007).
- 1505 51. Hazlett, J., Stamp, L.K., Merriman, T., Highton, J. & Hessian, P.A. IL-23R rs11209026  
1506 polymorphism modulates IL-17A expression in patients with rheumatoid arthritis. *Genes Immun*  
1507 **13**, 282-7 (2012).
- 1508 52. Karaderi, T. *et al.* Association between the interleukin 23 receptor and ankylosing spondylitis is  
1509 confirmed by a new UK case-control study and meta-analysis of published series. *Rheumatology*  
1510 (*Oxford*) **48**, 386-9 (2009).

- 1511 53. Duerr, R.H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel  
1512 disease gene. *Science* **314**, 1461-3 (2006).
- 1513 54. Abdollahi, E., Tavasolian, F., Momtazi-Borojeni, A.A., Samadi, M. & Rafatpanah, H. Protective  
1514 role of R381Q (rs11209026) polymorphism in IL-23R gene in immune-mediated diseases: A  
1515 comprehensive review. *J Immunotoxicol* **13**, 286-300 (2016).
- 1516 55. Abraham, C., Dulai, P.S., Vermeire, S. & Sandborn, W.J. Lessons Learned From Trials Targeting  
1517 Cytokine Pathways in Patients With Inflammatory Bowel Diseases. *Gastroenterology* **152**, 374-  
1518 388 e4 (2017).
- 1519 56. Molinelli, E., Campanati, A., Ganzetti, G. & Offidani, A. Biologic Therapy in Immune Mediated  
1520 Inflammatory Disease: Basic Science and Clinical Concepts. *Curr Drug Saf* **11**, 35-43 (2016).
- 1521 57. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-7 (2016).
- 1522 58. Wells, J.C. Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* **21**,  
1523 415-30 (2007).
- 1524 59. Loomba-Albrecht, L.A. & Styne, D.M. Effect of puberty on body composition. *Curr Opin*  
1525 *Endocrinol Diabetes Obes* **16**, 10-5 (2009).
- 1526 60. Rogol, A.D., Roemmich, J.N. & Clark, P.A. Growth at puberty. *J Adolesc Health* **31**, 192-200  
1527 (2002).
- 1528 61. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-45 (2012).
- 1529 62. Stern, J.H., Rutkowski, J.M. & Scherer, P.E. Adiponectin, Leptin, and Fatty Acids in the  
1530 Maintenance of Metabolic Homeostasis through Adipose Tissue Crosstalk. *Cell Metab* **23**, 770-84  
1531 (2016).
- 1532 63. Dewey, F.E. *et al.* Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J*  
1533 *Med* **374**, 1123-33 (2016).

- 1534 64. Bondestam, J. *et al.* cDNA cloning, expression studies and chromosome mapping of human type  
1535 I serine/threonine kinase receptor ALK7 (ACVR1C). *Cytogenet Cell Genet* **95**, 157-62 (2001).
- 1536 65. Jornvall, H., Blokzijl, A., ten Dijke, P. & Ibanez, C.F. The orphan receptor serine/threonine kinase  
1537 ALK7 signals arrest of proliferation and morphological differentiation in a neuronal cell line. *J*  
1538 *Biol Chem* **276**, 5140-6 (2001).
- 1539 66. Kim, B.C. *et al.* Activin receptor-like kinase-7 induces apoptosis through activation of MAPKs in a  
1540 Smad3-dependent mechanism in hepatoma cells. *J Biol Chem* **279**, 28458-65 (2004).
- 1541 67. Watanabe, R. *et al.* The MH1 domains of smad2 and smad3 are involved in the regulation of the  
1542 ALK7 signals. *Biochem Biophys Res Commun* **254**, 707-12 (1999).
- 1543 68. Kogame, M. *et al.* ALK7 is a novel marker for adipocyte differentiation. *J Med Invest* **53**, 238-45  
1544 (2006).
- 1545 69. Murakami, M. *et al.* Expression of activin receptor-like kinase 7 in adipose tissues. *Biochem*  
1546 *Genet* **51**, 202-10 (2013).
- 1547 70. Carlsson, L.M. *et al.* ALK7 expression is specific for adipose tissue, reduced in obesity and  
1548 correlates to factors implicated in metabolic disease. *Biochem Biophys Res Commun* **382**, 309-14  
1549 (2009).
- 1550 71. Carithers, L.J. & Moore, H.M. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv*  
1551 *Biobank* **13**, 307-8 (2015).
- 1552 72. Yogosawa, S., Mizutani, S., Ogawa, Y. & Izumi, T. Activin receptor-like kinase 7 suppresses  
1553 lipolysis to accumulate fat in obesity through downregulation of peroxisome proliferator-  
1554 activated receptor gamma and C/EBPalpha. *Diabetes* **62**, 115-23 (2013).
- 1555 73. Yogosawa, S. & Izumi, T. Roles of activin receptor-like kinase 7 signaling and its target,  
1556 peroxisome proliferator-activated receptor gamma, in lean and obese adipocytes. *Adipocyte* **2**,  
1557 246-50 (2013).



- 1558 74. Seifi, M., Ghasemi, A., Namipashaki, A. & Samadikuchaksaraei, A. Is C771G polymorphism of  
1559 MLX interacting protein-like (MLXIPL) gene a novel genetic risk factor for non-alcoholic fatty liver  
1560 disease? *Cell Mol Biol (Noisy-le-grand)* **60**, 37-42 (2014).
- 1561 75. Cairo, S., Merla, G., Urbinati, F., Ballabio, A. & Reymond, A. WBSR14, a gene mapping to the  
1562 Williams--Beuren syndrome deleted region, is a new member of the Mlx transcription factor  
1563 network. *Hum Mol Genet* **10**, 617-27 (2001).
- 1564 76. Ambele, M.A., Dessels, C., Durandt, C. & Pepper, M.S. Genome-wide analysis of gene expression  
1565 during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene  
1566 expression during adipocyte differentiation. *Stem Cell Res* **16**, 725-34 (2016).
- 1567 77. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet* **46**, 200-4  
1568 (2014).
- 1569 78. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and  
1570 population analysis. *Bioinformatics* **28**, 2543-5 (2012).
- 1571 79. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat*  
1572 *Protoc* **9**, 1192-212 (2014).
- 1573 80. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution.  
1574 *Nature* **518**, 187-196 (2015).
- 1575 81. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**,  
1576 185-90 (2014).
- 1577 82. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12  
1578 (2011).
- 1579 83. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies  
1580 additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).

- 1581 84. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range  
1582 of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
- 1583 85. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for  
1584 genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
- 1585 86. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven  
1586 common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
- 1587 87. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev*  
1588 *Genet* **11**, 499-511 (2010).
- 1589 88. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6  
1590 (2007).
- 1591 89. Moayyeri, A., Hammond, C.J., Valdes, A.M. & Spector, T.D. Cohort Profile: TwinsUK and healthy  
1592 ageing twin study. *Int J Epidemiol* **42**, 76-85 (2013).
- 1593 90. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon  
1594 Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-27 (2013).
- 1595 91. Kutalik, Z., Whittaker, J., Waterworth, D., Beckmann, J.S. & Bergmann, S. Novel method to  
1596 estimate the phenotypic variation explained by genome-wide association studies reveals large  
1597 fraction of the missing heritability. *Genet Epidemiol* **35**, 341-9 (2011).
- 1598 92. Billingsley, P. *Probability and measure*, xii, 622 p. (Wiley, New York, 1986).
- 1599 93. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated  
1600 with blood pressure and hypertension. *Nat Genet* **48**, 1151-61 (2016).
- 1601 94. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-  
1602 analysis of coronary artery disease. *Nat Genet* **47**, 1121-30 (2015).
- 1603 95. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U*  
1604 *S A* **100**, 9440-5 (2003).

- 1605 96. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits.  
1606 *Am J Hum Genet* **100**, 428-443 (2017).
- 1607 97. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**, D222-6  
1608 (2015).
- 1609 98. Toyofuku, T. *et al.* Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses  
1610 angiogenesis via Plexin-D1. *EMBO J* **26**, 1373-84 (2007).
- 1611 99. Gitler, A.D., Lu, M.M. & Epstein, J.A. PlexinD1 and semaphorin signaling are required in  
1612 endothelial cells for cardiovascular development. *Dev Cell* **7**, 107-16 (2004).
- 1613 100. Luchino, J. *et al.* Semaphorin 3E suppresses tumor cell death triggered by the plexin D1  
1614 dependence receptor in metastatic breast cancers. *Cancer Cell* **24**, 673-85 (2013).
- 1615 101. Shimizu, I. *et al.* Semaphorin3E-induced inflammation contributes to insulin resistance in dietary  
1616 obesity. *Cell Metab* **18**, 491-504 (2013).
- 1617 102. Verzijl, H.T., van der Zwaag, B., Cruysberg, J.R. & Padberg, G.W. Mobius syndrome redefined: a  
1618 syndrome of rhombencephalic maldevelopment. *Neurology* **61**, 327-33 (2003).
- 1619 103. Verzijl, H.T., van der Zwaag, B., Lammens, M., ten Donkelaar, H.J. & Padberg, G.W. The  
1620 neuropathology of hereditary congenital facial palsy vs Mobius syndrome. *Neurology* **64**, 649-53  
1621 (2005).
- 1622 104. Fujita, M., Reinhart, F. & Neutra, M. Convergence of apical and basolateral endocytic pathways  
1623 at apical late endosomes in absorptive cells of suckling rat ileum in vivo. *J Cell Sci* **97 ( Pt 2)**, 385-  
1624 94 (1990).
- 1625 105. Briegel, W. Neuropsychiatric findings of Mobius sequence -- a review. *Clin Genet* **70**, 91-7 (2006).
- 1626 106. Ta-Shma, A. *et al.* Isolated truncus arteriosus associated with a mutation in the plexin-D1 gene.  
1627 *Am J Med Genet A* **161A**, 3115-20 (2013).

- 1628 107. Mazzotta, C. *et al.* Plexin-D1/Semaphorin 3E pathway may contribute to dysregulation of  
1629 vascular tone control and defective angiogenesis in systemic sclerosis. *Arthritis Res Ther* **17**, 221  
1630 (2015).
- 1631 108. Yang, W.J. *et al.* Semaphorin-3C signals through Neuropilin-1 and PlexinD1 receptors to inhibit  
1632 pathological angiogenesis. *EMBO Mol Med* **7**, 1267-84 (2015).
- 1633 109. Zygmunt, T. *et al.* Semaphorin-PlexinD1 signaling limits angiogenic potential via the VEGF decoy  
1634 receptor sFlt1. *Dev Cell* **21**, 301-14 (2011).
- 1635 110. Kim, J., Oh, W.J., Gaiano, N., Yoshida, Y. & Gu, C. Semaphorin 3E-Plexin-D1 signaling regulates  
1636 VEGF function in developmental angiogenesis via a feedback mechanism. *Genes Dev* **25**, 1399-  
1637 411 (2011).
- 1638 111. Bertolino, P. *et al.* Activin B receptor ALK7 is a negative regulator of pancreatic beta-cell  
1639 function. *Proc Natl Acad Sci U S A* **105**, 7246-51 (2008).
- 1640 112. Haworth, K.E. *et al.* Methylation of the FGFR2 gene is associated with high birth weight centile in  
1641 humans. *Epigenomics* **6**, 477-91 (2014).
- 1642 113. Chi, X. *et al.* Angiopoietin-like 4 Modifies the Interactions between Lipoprotein Lipase and Its  
1643 Endothelial Cell Transporter GPIHBP1. *J Biol Chem* **290**, 11865-77 (2015).
- 1644 114. Catoire, M. *et al.* Fatty acid-inducible ANGPTL4 governs lipid metabolic response to exercise.  
1645 *Proc Natl Acad Sci U S A* **111**, E1043-52 (2014).
- 1646 115. van Raalte, D.H. *et al.* Angiopoietin-like protein 4 is differentially regulated by glucocorticoids  
1647 and insulin in vitro and in vivo in healthy humans. *Exp Clin Endocrinol Diabetes* **120**, 598-603  
1648 (2012).
- 1649 116. Koster, A. *et al.* Transgenic angiopoietin-like (angptl)4 overexpression and targeted disruption of  
1650 angptl4 and angptl3: regulation of triglyceride metabolism. *Endocrinology* **146**, 4943-50 (2005).

- 1651 117. Thiagalingam, A. *et al.* RREB-1, a novel zinc finger protein, is involved in the differentiation  
1652 response to Ras in human medullary thyroid carcinomas. *Mol Cell Biol* **16**, 5335-45 (1996).
- 1653 118. Bonomo, J.A. *et al.* The ras responsive transcription factor RREB1 is a novel candidate gene for  
1654 type 2 diabetes associated end-stage kidney disease. *Hum Mol Genet* **23**, 6441-7 (2014).
- 1655 119. Thiagalingam, A., Lengauer, C., Baylin, S.B. & Nelkin, B.D. RREB1, a ras responsive element  
1656 binding protein, maps to human chromosome 6p25. *Genomics* **45**, 630-2 (1997).
- 1657 120. Bisogno, T. *et al.* Cloning of the first sn1-DAG lipases points to the spatial and temporal  
1658 regulation of endocannabinoid signaling in the brain. *J Cell Biol* **163**, 463-8 (2003).
- 1659 121. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat*  
1660 *Genet* **45**, 1274-83 (2013).
- 1661 122. Kooner, J.S. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma  
1662 triglycerides. *Nat Genet* **40**, 149-51 (2008).
- 1663 123. Pan, L.A. *et al.* G771C Polymorphism in the MLXIPL Gene Is Associated with a Risk of Coronary  
1664 Artery Disease in the Chinese: A Case-Control Study. *Cardiology* **114**, 174-8 (2009).
- 1665 124. Kang, G., Leech, C.A., Chepurny, O.G., Coetzee, W.A. & Holz, G.G. Role of the cAMP sensor Epac  
1666 as a determinant of KATP channel ATP sensitivity in human pancreatic beta-cells and rat INS-1  
1667 cells. *J Physiol* **586**, 1307-19 (2008).
- 1668 125. Ji, Z., Mei, F.C. & Cheng, X. Epac, not PKA catalytic subunit, is required for 3T3-L1 preadipocyte  
1669 differentiation. *Front Biosci (Elite Ed)* **2**, 392-8 (2010).
- 1670 126. Martini, C.N., Plaza, M.V. & Vila Mdel, C. PKA-dependent and independent cAMP signaling in  
1671 3T3-L1 fibroblasts differentiation. *Mol Cell Endocrinol* **298**, 42-7 (2009).
- 1672 127. Petersen, R.K. *et al.* Cyclic AMP (cAMP)-mediated stimulation of adipocyte differentiation  
1673 requires the synergistic action of Epac- and cAMP-dependent protein kinase-dependent  
1674 processes. *Mol Cell Biol* **28**, 3804-16 (2008).

- 1675 128. Yan, J. *et al.* Enhanced leptin sensitivity, reduced adiposity, and improved glucose homeostasis  
1676 in mice lacking exchange protein directly activated by cyclic AMP isoform 1. *Mol Cell Biol* **33**,  
1677 918-26 (2013).
- 1678 129. Gesta, S. *et al.* Evidence for a role of developmental genes in the origin of obesity and body fat  
1679 distribution. *Proc Natl Acad Sci U S A* **103**, 6676-81 (2006).
- 1680 130. Gesta, S. *et al.* Mesodermal developmental gene Tbx15 impairs adipocyte differentiation and  
1681 mitochondrial respiration. *Proc Natl Acad Sci U S A* **108**, 2771-6 (2011).
- 1682 131. Lee, K.Y. *et al.* Tbx15 controls skeletal muscle fibre-type determination and muscle metabolism.  
1683 *Nat Commun* **6**, 8054 (2015).
- 1684

## 36 TABLES

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

Locus (+/- 1Mb of a given variant)	Chr:Position (GRCh37) <sup>b</sup>	rsID	EA	OA	Gene <sup>c</sup>	Amino Acid Change <sup>e</sup>	If locus is known, nearby (< 1 MB) published variant(s) <sup>d</sup>	N	EAF	$\beta^g$	SE	P-value	P-value for Sex-heterogeneity <sup>f</sup>	Other Criteria For Sig <sup>h</sup>
<b>Variants in Novel Loci</b>														
<b>All Ancestry Additive model Sex-combined analyses</b>														
1	2:158412701	rs55920843	T	G	<i>ACVR1C</i>	N150H	-	455,526	0.989	0.065	0.011	<b>4.8E-10</b>	<b>1.7E-07</b>	
2	3:50597092	rs1034405	G	A	<i>C3orf18</i>	A162V	-	455,424	0.135	0.016	0.003	<b>1.9E-07</b>	8.8E-01	G,C
3	4:120528327	rs3733526	G	A	<i>PDE5A</i>	A41V	-	461,521	0.187	0.015	0.003	<b>2.6E-08</b>	5.2E-03	
4	6:26108117	rs146860658	T	C	<i>HIST1H1T</i>	A69T	-	217,995	0.001	0.229	0.042	<b>4.3E-08</b>	6.3E-01	S
5	7:6449496	rs2303361	C	T	<i>DAGLB</i>	Q664R	-	475,748	0.221	0.014	0.003	<b>6.2E-08</b>	3.4E-03	G
6	10:123279643	rs138315382	T	C	<i>FGFR2</i>	synonymous	-	236,962	0.001	0.258	0.049	<b>1.4E-07</b>	1.1E-01	G,S
7	11:65403651	rs7114037	C	A	<i>PCNXL3</i>	H1822Q	-	448,861	0.954	0.029	0.005	<b>1.8E-08</b>	4.4E-01	
8	12:48143315	rs145878042	A	G	<i>RAPGEF3</i>	L300P	-	470,513	0.990	0.085	0.010	<b>7.2E-17</b>	7.3E-03	
9	12:108618630	rs3764002	C	T	<i>WSCD2</i>	T266I	-	474,637	0.737	0.014	0.002	<b>9.8E-10</b>	5.5E-01	
10	15:42032383	rs17677991	G	C	<i>MGA</i>	P1523A	-	469,874	0.345	0.015	0.002	<b>3.5E-11</b>	9.1E-01	
11	16:4432029	rs3810818	A	C	<i>VASN</i>	E384A	-	424,163	0.231	0.016	0.003	<b>2.0E-09</b>	3.3E-01	
	16:4445327	rs3747579	C	T	<i>CORO7</i>	R193Q	-	453,078	0.299	0.018	0.002	<b>2.2E-13</b>	4.3E-02	
12	16:4484396	rs1139653	A	T	<i>DNAJA3</i>	N75Y	-	434,331	0.284	0.015	0.002	<b>4.3E-10</b>	1.4E-01	
	19:49232226	rs2287922	A	G	<i>RASIP1</i>	R601C	-	430,272	0.494	0.014	0.002	<b>1.6E-09</b>	3.7E-02	
13	19:49244220	rs2307019	G	A	<i>IZUMO1</i>	A333V	-	476,147	0.558	0.012	0.002	<b>4.7E-08</b>	3.9E-02	
	20:42965811	rs144098855	T	C	<i>R3HDML</i>	P5L	-	428,768	0.001	0.172	0.032	<b>9.7E-08</b>	1.0E+00	G
<b>European Ancestry Additive model Sex-combined analyses</b>														
14	1:173802608	rs35515638	G	A	<i>DARS2</i>	K196R	-	352,646	0.001	0.201	0.038	<b>1.4E-07</b>	6.0E-02	G
15	14:58838668	rs1051860	A	G	<i>ARID4A</i>	synonymous	-	367,079	0.411	0.013	0.002	<b>2.2E-08</b>	1.3E-01	
16	15:42115747	rs3959569	C	G	<i>MAPKBP1</i>	R1240H	-	253,703	0.349	0.017	0.003	<b>2.0E-08</b>	6.3E-01	
<b>Variants in Previously Identified Loci</b>														
<b>All Ancestry Additive model Sex-combined analyses</b>														
1	1:119427467	rs61730011	A	C	<i>TBX15</i>	M566R	rs2645294, rs12731372, rs12143789, rs1106529	441,461	0.957	0.041	0.005	<b>2.2E-14</b>	6.7E-01	
	1:119469188	rs10494217	T	G		H156N	472,259	0.174	0.018	0.003	<b>1.4E-10</b>	6.0E-01		
2	1:154987704	rs141845046	C	T	<i>ZBTB7B</i>	P190S	rs905938	476,440	0.976	0.037	0.007	<b>3.8E-08</b>	<b>7.9E-07</b>	C

bioRxiv preprint doi: <https://doi.org/10.1101/352674>; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3	2:165551201	rs7607980	T	C	<i>COBLL1</i>	N941D	rs1128249, rs10195252, rs12692737, rs12692738, rs17185198	389,883	0.879	0.026	0.004	<b>1.6E-13</b>	3.0E-30	
4	2:188343497	rs7586970	T	C	<i>TFPI</i>	N221S	rs1569135	452,638	0.697	0.016	0.002	<b>3.0E-12</b>	6.3E-01	
5	3:52558008	rs13303	T	C	<i>STAB1</i>	M113T	rs2276824	470,111	0.445	0.019	0.002	<b>5.5E-18</b>	6.7E-02	
	3:52833805	rs3617	C	A	<i>ITIH3</i>	Q315K		452,150	0.541	0.015	0.002	<b>1.6E-12</b>	4.0E-01	C
6	3:129137188	rs62266958	C	T	<i>EFCAB12</i>	R197H	rs10804591	476,382	0.936	0.036	0.004	<b>8.3E-17</b>	<b>9.3E-05</b>	
	3:129284818	rs2625973	A	C	<i>PLXND1</i>	L1412V		476,338	0.733	0.016	0.002	<b>9.2E-11</b>	<b>1.6E-05</b>	
7	4:89625427	rs1804080	G	C	<i>HERC3</i>	E946Q	rs9991328	446,080	0.838	0.021	0.003	<b>1.5E-12</b>	<b>4.1E-06</b>	
	4:89668859	rs7657817	C	T	<i>FAM13A</i>	V443I		476,383	0.815	0.016	0.003	<b>5.0E-09</b>	<b>9.6E-05</b>	
8	5:176516631	rs1966265	A	G	<i>FGFR4</i>	V10I	rs6556301	455,246	0.236	0.023	0.003	<b>1.7E-19</b>	2.1E-01	
9	6:7211818	<b>rs1334576<sup>E</sup></b>	G	A	<i>RREB1</i>	G195R	rs1294410	451,044	0.565	0.017	0.002	<b>3.9E-15</b>	1.5E-01	
10	6:34827085	rs9469913	A	T	<i>UHRF1BP1</i>	Q984H	rs1776897	309,684	0.847	0.021	0.004	<b>1.2E-08</b>	2.7E-01	C
11	6:127476516	rs1892172	A	G	<i>RSP03</i>	synonymous	rs11961815, rs72959041, rs1936805	476,358	0.543	0.031	0.002	<b>2.6E-47</b>	<b>7.7E-09</b>	
	6:127767954	<b>rs139745911<sup>E</sup></b>	A	G	<i>KIAA0408</i>	P504S		391,469	0.010	0.103	0.012	<b>6.8E-19</b>	<b>2.0E-04</b>	
12	7:73012042	rs35332062	G	A	<i>MLXIPL</i>	A358V	rs6976930	451,158	0.880	0.020	0.003	<b>1.8E-09</b>	1.5E-01	
	7:73020337	rs3812316	C	G		Q241H		454,738	0.881	0.021	0.003	<b>2.0E-10</b>	5.8E-02	
13	10:95931087	rs17417407	T	G	<i>PLCE1</i>	R240L	rs10786152	476,475	0.173	0.018	0.003	<b>2.5E-11</b>	5.9E-01	
14	11:64031241	rs35169799	T	C	<i>PLCB3</i>	S778L	rs11231693	476,457	0.061	0.034	0.004	<b>9.1E-15</b>	<b>1.3E-04</b>	
15	12:123444507	rs58843120	G	T	<i>DNAH10</i>	F92L	rs4765219, rs863750	466,498	0.987	0.053	0.009	<b>1.3E-08</b>	3.5E-01	
	12:124265687	rs11057353	T	C		S228P		476,360	0.373	0.018	0.002	<b>2.1E-16</b>	<b>2.7E-08</b>	
	12:124330311	rs34934281	C	T		T1785M		476,395	0.889	0.025	0.003	<b>2.9E-14</b>	<b>3.1E-08</b>	
	12:124427306	rs11057401	T	A		<i>CCDC92</i>		S53C	467,649	0.695	0.029	0.002	<b>7.3E-37</b>	<b>5.5E-11</b>
16	15:56756285	rs1715919	G	T	<i>MNS1</i>	Q55P	rs8030605	476,274	0.096	0.023	0.004	<b>8.8E-11</b>	2.7E-02	
17	16:67397580	rs9922085	G	C	<i>LRRC36</i>	R101P	rs6499129	469,474	0.938	0.034	0.005	<b>3.8E-13</b>	5.9E-01	
	16:67409180	rs8052655	G	A		G388S		474,035	0.939	0.034	0.005	<b>5.5E-13</b>	4.0E-01	
18	19:18285944	rs11554159	A	G	<i>IFI30</i>	R76Q	rs12608504	476,389	0.257	0.015	0.002	<b>3.5E-10</b>	3.1E-03	
	19:18304700	rs874628	G	A	<i>MPV17L2</i>	M72V		476,388	0.271	0.015	0.002	<b>1.2E-10</b>	2.5E-03	
19	20:33971914	rs4911494	T	C	<i>UQCC1</i>	R51Q	rs224333	451,064	0.602	0.018	0.002	<b>2.5E-16</b>	1.5E-03	
	20:34022387	rs224331	A	C	<i>GDF5</i>	S276A		345,805	0.644	0.017	0.003	<b>1.8E-11</b>	3.2E-03	
<b>All Ancestry Recessive model Sex-combined analyses</b>														
20	17:17425631	rs897453	C	T	<i>PEMT</i>	V58L	rs4646404	476,546	0.569	0.025	0.004	<b>4.1E-11</b>	8.2E-01	
<b>European Ancestry Additive model Sex-combined analyses</b>														
6	3:129293256	rs2255703	T	C	<i>PLXND1</i>	M870V	rs10804591	420,520	0.620	0.014	0.002	<b>3.1E-09</b>	<b>1.6E-04</b>	

bioRxiv preprint doi: <https://doi.org/10.1101/352674>; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: <https://doi.org/10.1101/352674>; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

39 Abbreviations: GRCh37=human genome assembly build37;rsID=based on dbSNP; VEP=Ensembl Variant Effect Predictor toolset; GTEx=Genotype-Tissue Expression project;SD=standard deviation; SE=standard error;N=sample size; EAF=effect allele  
40 frequency; EA=effect allele; OA=other allele.  
41 a Coding variants refer to variants located in the exons and splicing junction regions.  
42 b Variant positions are reported according to Human assembly build 37 and their alleles are coded based on the positive strand.  
43 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).  
44 d Previously published variants within +/-1Mb are from Shungin et al.<sup>10</sup>, except for rs6976930 and rs10786152 from Graff et al.<sup>14</sup> and rs6499129 from Ng. et al.<sup>16</sup>.  
45 e Effect size is based on standard deviation (SD) per effect allele  
46 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 stratified genome-wide  
47 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).  
48 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, rs1936805, in a  
49 known locus (see Supplementary 8A/B).  
50 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 of Stage 1  
51 association.  
52

03  
04

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

Locus (+/-1Mb of a given variant)	Chr:Position (GRCh37) <sup>c</sup>	rsID	EA	OA	Gene <sup>d</sup>	Amino Acid Change <sup>d</sup>	In sex-combined analyses <sup>e</sup>	If locus is known, nearby (< 1 MB) published variant(s) <sup>f</sup>	P-value for Sex-heterogeneity <sup>g</sup>	Men					Women					Other Criteria For Sig <sup>j</sup>
										N	EAF	$\beta^h$	SE	P	N	EAF	$\beta^h$	SE	P	
<b>Variants in Novel Loci</b>																				
<b>All Ancestry Additive model Men only analyses</b>																				
1	13:96665697	rs148108950	A	G	<i>UGGT2</i>	P175L	No	-	<b>1.5E-06</b>	203,009	0.006	0.130	0.024	<b>6.0E-08</b>	221,390	0.004	-0.044	0.027	1.1E-01	G
2	14:23312594	rs1042704	A	G	<i>MMP14</i>	D273N	No	-	<b>2.6E-04</b>	226,646	0.202	0.021	0.004	<b>2.0E-08</b>	250,018	0.197	0.002	0.004	6.1E-01	
<b>All Ancestry Additive model Women only analyses</b>																				
3	1:205130413	rs3851294	G	A	<i>DSTYK</i>	C641R	No	-	<b>9.8E-08</b>	225,803	0.914	-0.005	0.005	<b>3.0E-11</b>	249,471	0.912	0.034	0.005	<b>4.5E-11</b>	
4	2:158412701	rs55920843	T	G	<i>ACVR1C</i>	N150H	Yes	-	<b>1.7E-07</b>	210,071	0.989	0.006	0.015	<b>7.0E-11</b>	245,808	0.989	0.113	0.014	<b>1.7E-15</b>	
5	19:8429323	rs116843064	G	A	<i>ANGPTL4</i>	E40K	No	-	<b>1.3E-07</b>	203,098	0.981	-0.017	0.011	<b>1.0E-11</b>	243,351	0.981	0.064	0.011	<b>1.2E-09</b>	
<b>Variants in Previously Identified Loci</b>																				
<b>All Ancestry Additive model Women only analyses</b>																				
1	1:154987704	rs141845046	C	T	<i>ZBTB7B</i>	P190S	Yes	rs905938	<b>7.9E-07</b>	226,709	0.975	0.004	0.010	<b>6.9E-11</b>	250,084	0.977	0.070	0.010	<b>2.3E-13</b>	
2	2:165551201	rs7607980	T	C	<i>COBLL1</i>	N941D	Yes	rs1128249, rs10195252, rs12692737, rs12692738, rs17185198	<b>3.0E-30</b>	173,600	0.880	-0.018	0.005	<b>5.0E-04</b>	216,636	0.878	0.062	0.005	<b>6.7E-39</b>	
3	3:129137188	rs62266958	C	T	<i>EFCAB12</i>	R197H	Yes	rs10804591	<b>9.3E-05</b>	226,690	0.937	0.018	0.006	<b>3.0E-03</b>	250,045	0.936	0.051	0.006	<b>8.1E-18</b>	
	3:129284818	rs2625973	A	C	<i>PLXND1</i>	L1412V	Yes		<b>1.6E-05</b>	226,650	0.736	0.005	0.003	<b>1.0E-08</b>	250,023	0.730	0.025	0.003	<b>8.2E-14</b>	
	3:129293256	rs2255703	T	C		M870V	Yes		<b>5.0E-04</b>	226,681	0.609	0.003	0.003	<b>3.0E-11</b>	250,069	0.602	0.018	0.003	<b>1.9E-09</b>	
4	4:89625427	rs1804080	G	C	<i>HERC3</i>	E946Q	Yes	rs9991328	<b>4.1E-06</b>	222,556	0.839	0.008	0.004	<b>6.0E-02</b>	223,877	0.837	0.034	0.004	<b>2.1E-16</b>	
	4:89668859	rs7657817	C	T	<i>FAM13A</i>	V443I	Yes		<b>9.6E-05</b>	226,680	0.816	0.006	0.004	<b>1.0E-11</b>	242,970	0.815	0.026	0.004	<b>5.9E-12</b>	
5	6:127476516	rs1892172	A	G	<i>RSPO3</i>	synonymous	Yes	rs11961815, rs72959041, rs1936805	<b>7.7E-09</b>	226,677	0.541	0.018	0.003	<b>5.0E-10</b>	250,034	0.545	0.042	0.003	<b>3.4E-48</b>	
	6:127767954	rs139745911	A	G	<i>KIAA0408</i>	P504S	Yes		<b>2.0E-04</b>	188,079	0.010	0.057	0.017	<b>6.0E-04</b>	205,203	0.010	0.143	0.016	<b>5.9E-19</b>	
6	11:64031241	rs35169799	T	C	<i>PLCB3</i>	S778L	Yes	rs11231693	<b>1.3E-04</b>	226,713	0.061	0.016	0.006	<b>9.0E-03</b>	250,097	0.061	0.049	0.006	<b>6.7E-16</b>	
7	12:124265687	rs11057353	T	C	<i>DNAH10</i>	S228P	Yes	rs4765219, rs863750	<b>2.7E-08</b>	226,659	0.370	0.005	0.003	<b>8.0E-12</b>	250,054	0.376	0.029	0.003	<b>3.1E-22</b>	
	12:124330311	rs34934281	C	T		T1785M	Yes		<b>3.1E-08</b>	226,682	0.891	0.006	0.005	<b>1.9E-11</b>	250,066	0.887	0.043	0.005	<b>1.4E-20</b>	
	12:124427306	rs11057401	T	A	<i>CCDC92</i>	S53C	Yes		<b>5.5E-11</b>	223,324	0.701	0.013	0.003	<b>4.3E-05</b>	244,678	0.689	0.043	0.003	<b>1.0E-41</b>	

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 allele; OA=other allele; EAF=effect allele frequency.  
a Coding variants refer to variants located in the exons and splicing junction regions.  
b Bonferonni corrected Pvalue for the number of SNPs tested for sex-heterogeneity is <7.14E-04 i.e. 0.05/70 variants.  
c Variant positions are reported according to Human assembly build 37 and their alleles are coded based on the positive strand.

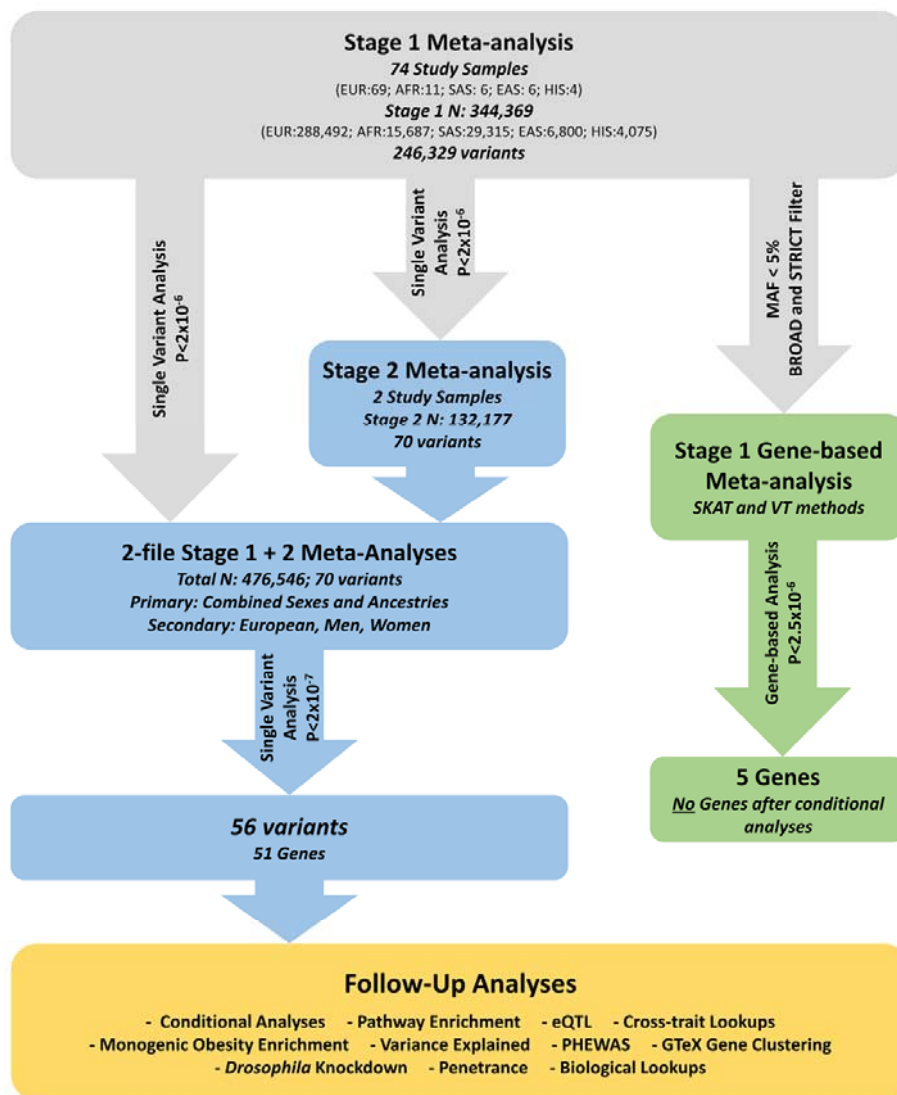
bioRxiv preprint doi: <https://doi.org/10.1101/352074>; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

- L0 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).
- L1 e Variant was also identified as array-wide significant in the sex-combined analyses.
- L2 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).
- L3 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 genome-wide association meta-analysis data. Bioinformatics 2015; 31, 259-61. PMID: 25260699.
- L4 h Effect size is based on standard deviation (SD) per effect allele
- L5 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).
- L6 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 replicable in Stage 2 studies for validation of Stage 1 association.
- L7
- L8
- L9

1 **FIGURES**

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

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



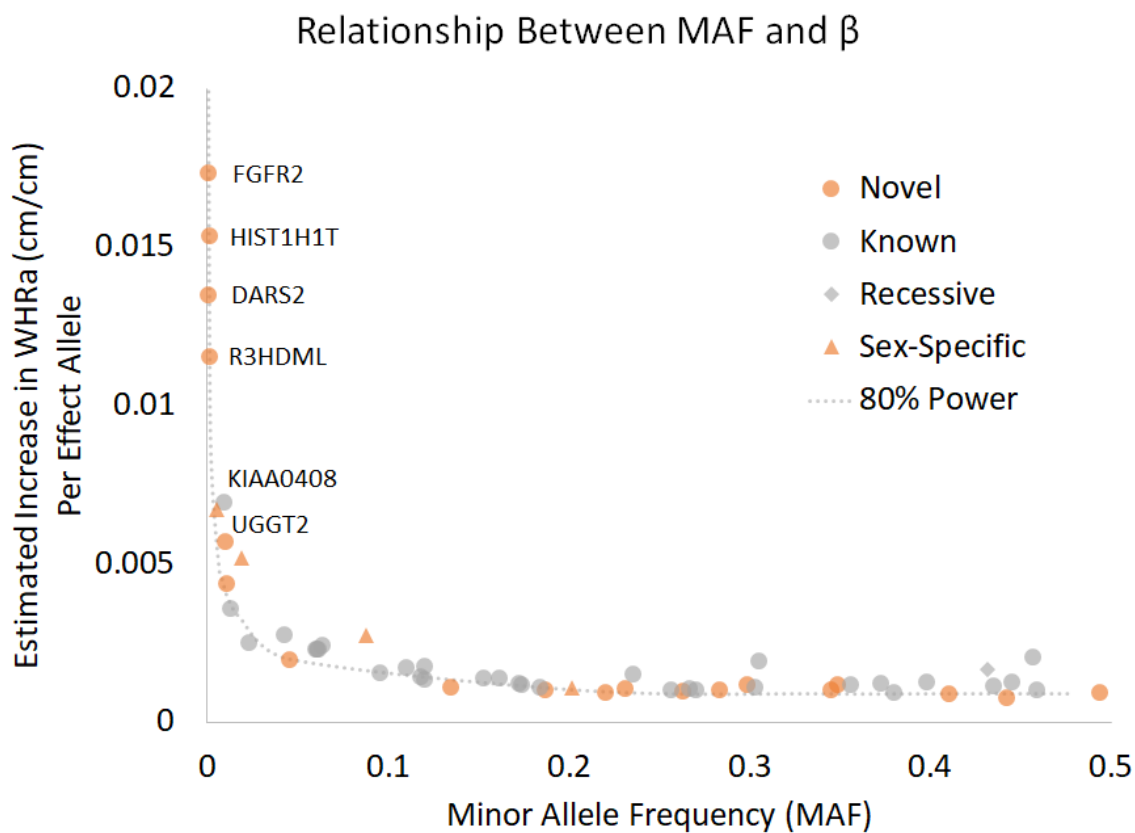
**Number of Significant Variants (All Ancestries/European)**

Strata	ADD	REC	NOVEL*	GWAS INDEP <sup>Y</sup>
Combined	47/4	1/0	16/3	1/0
Women	16/0	0/0	3/0	0/0
Men	2/0	0/0	2/0	0/0
<b>Total Non-Overlapping Signals</b>	<b>56 Variants Across 41 Signals</b>		<b>24 New Signals</b>	

\*Novel variants include those that are >1MB from a previously published WHRadjBMI GWAS tag SNP.

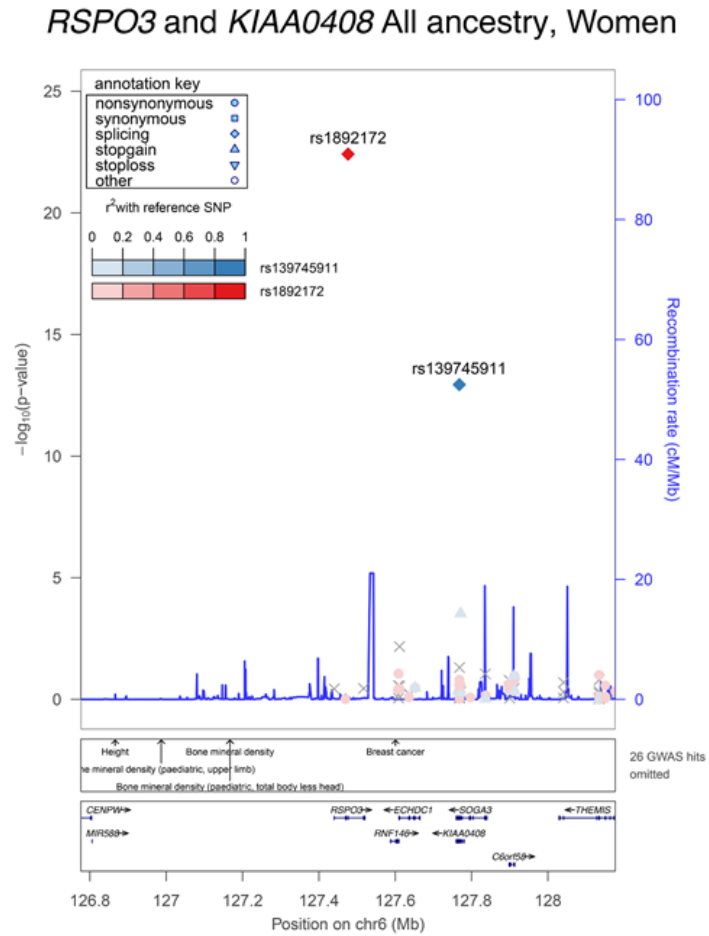
<sup>Y</sup> Independent (INDEP) includes variants that are nearby known WHRadjBMI GWAS tag variants, but were determined independent after conditional analysis.

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

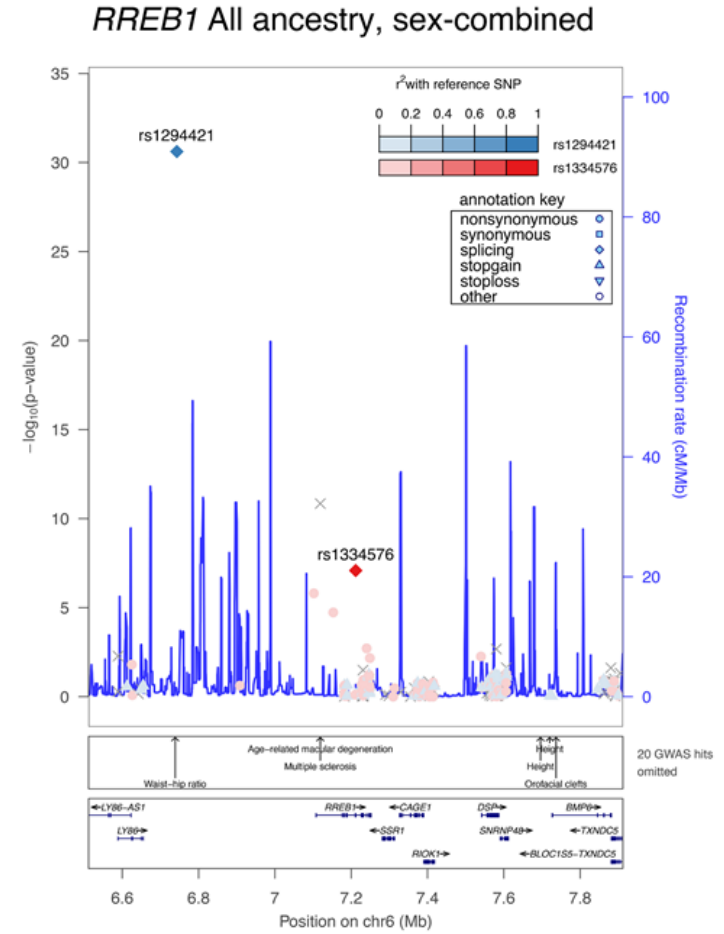


14 **Figure 3.** Regional association plots for known loci with novel coding signals. Point color reflects  $r^2$  calculated from the ARIC dataset. In a) there  
 15 are two independent variants in *RSPO3* and *KIAA0408*, as shown by conditional analysis. In b) we have a variant in *RREB1* that is independent of  
 16 the GWAS variant rs1294421.  
 17

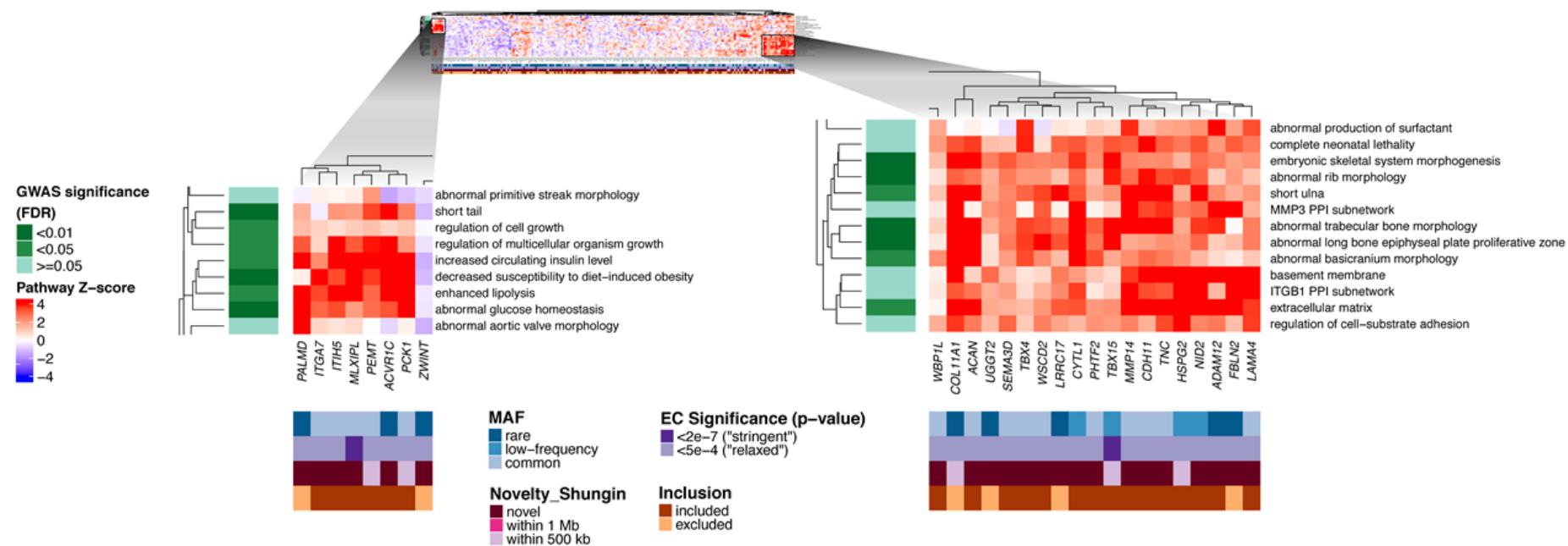
a)



b)



19 **Figure 4.** Heat maps showing DEPICT gene set enrichment results. For any given square, the color indicates how strongly the corresponding gene (shown on the x-axis) is predicted to belong to the reconstituted  
 20 gene set (y-axis). This value is based on the gene's z-score for gene set inclusion in DEPICT's reconstituted gene sets, where red indicates a higher and blue a lower z-score. To visually reduce redundancy and  
 21 increase clarity, we chose one representative "meta-gene set" for each group of highly correlated gene sets based on affinity propagation clustering (**Online Methods, Supplementary Note 2**). Heatmap  
 22 intensity and DEPICT P-values (see P-values in **Supplementary Data 4-5**) correspond to the most significantly enriched gene set within the meta-gene set. Annotations for the genes indicate (1) the minor allele  
 23 frequency of the significant ExomeChip (EC) variant (shades of blue; if multiple variants, the lowest-frequency variant was kept), (2) whether the variant's P-value reached array-wide significance ( $<2 \times 10^{-7}$ ) or  
 24 suggestive significance ( $<5 \times 10^{-4}$ ) (shades of purple), (3) whether the variant was novel, overlapping "relaxed" GWAS signals from Shungin et al.<sup>10</sup> (GWAS  $P < 5 \times 10^{-4}$ ) or overlapping "stringent" GWAS signals  
 25 (GWAS  $P < 5 \times 10^{-8}$ ) (shades of pink), and (4) whether the gene was included in the gene set enrichment analysis or excluded by filters (shades of brown/orange) (Online Methods and Supplementary  
 26 Information). Annotations for the gene sets indicate if the meta-gene set was found significant (shades of green; FDR  $< 0.01$ ,  $< 0.05$ , or not significant) in the DEPICT analysis of GWAS results from Shungin et al.  
 27  
 28



bioRxiv preprint doi: <https://doi.org/10.1101/352674>; this version posted June 30, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.