Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure underpinning obesity

Valérie Turcot[1*], Yingchang Lu[2,3,4*], Heather M Highland[5,6*], Claudia Schurmann[3,4*], Anne E Justice[5*], Rebecca S Fine[7,8,9*], Jonathan P Bradfield[10,11], Tõnu Esko[7,9,12], Ayush Giri[13], Mariaelisa Graff[5], Xiuqing Guo[14], Audrey E Hendricks[15,16], Tugce Karaderi[17,18], Adelheid Lempradl[19], Adam E Locke[20,21], Anubha Mahajan[17], Eirini Marouli[22], Suthesh Sivapalaratnam[23,24,25], Kristin L Young[5], Tamuno Alfred[3], Mary F Feitosa[26], Nicholas GD Masca[27,28], Alisa K Manning[7,24,29,30], Carolina Medina-Gomez[31,32], Poorva Mudgal[33], Maggie CY Ng[33,34], Alex P Reiner[35,36], Sailaja Vedantam[7,8,9], Sara M Willems[37], Thomas W Winkler[38], Goncalo Abecasis[20], Katja K Aben[39,40], Dewan S Alam[41], Sameer E Alharthi[22,42], Matthew Allison[43], Philippe Amouyel[44,45,46], Folkert W Asselbergs[47,48,49], Paul L Auer[50], Beverley Balkau[51], Lia E Bang[52], Inês Barroso[15,53,54], Lisa Bastarache[55], Marianne Benn[56,57], Sven Bergmann[58,59], Lawrence F Bielak[60], Matthias Blüher[61,62], Michael Boehnke[20], Heiner Boeing[63], Eric Boerwinkle[64,65], Carsten A Böger[66], Jette Bork-Jensen[67], Michiel L Bots[68], Erwin P Bottinger[3], Donald W Bowden[33,34,69], Ivan Brandslund[70,71], Gerome Breen[72,73], Murray H Brilliant[74], Linda Broer[32], Marco Brumat[75], Amber A Burt[76], Adam S Butterworth[77,78], Peter T Campbell[79], Stefania Cappellani[80], David J Carey[81], Eulalia Catamo[80], Mark J Caulfield[22,82], John C Chambers[83,84,85], Daniel I Chasman[7,86,87,88], Yii-Der Ida Chen[14], Rajiv Chowdhury[77], Cramer Christensen[89], Audrey Y Chu[87,90], Massimiliano Cocca[91], Francis S Collins[92], James P Cook[93], Janie Corley[94,95], Jordi Corominas Galbany[96,97], Amanda J Cox[33,34,98], David S Crosslin[99], Gabriel Cuellar-Partida[100,101], Angela D'Eustacchio[80], John Danesh[15,77,78,102], Gail Davies[94,95], Paul IW de Bakker[68,103], Mark CH de Groot[104,105], Renée de Mutsert[106], Ian J Deary[94,95], George Dedoussis[107], Ellen W Demerath[108], Martin den Heijer[109], Anneke I den Hollander[97], Hester M den Ruijter[110], Joe G Dennis[111], Josh C Denny[55], Emanuele Di Angelantonio[77,78], Fotios Drenos[112,113], Mengmeng Du[114,115], Marie-Pierre Dubé[1,116], Alison M Dunning[117], Douglas F Easton[111,117], Todd L Edwards[13], David Ellinghaus[118], Patrick T Ellinor[7,24,30], Paul Elliott[119], Evangelos Evangelou[84,120], Aliki-Eleni Farmaki[107,121], I. Sadaf Farooqi[53,54], Jessica D Faul[122], Sascha Fauser[123], Shuang Feng[20], Ele Ferrannini[124,125], Jean Ferrieres[126], Jose C Florez[7,24,29,30], Ian Ford[127], Myriam Fornage[128], Oscar H Franco[31], Andre Franke[118], Paul W Franks[129,130,131], Nele Friedrich[132], Ruth Frikke-Schmidt[57,133], Tessel E. Galesloot[40], Wei Gan[17], Ilaria Gandin[134], Paolo Gasparini[75,135], Jane Gibson[136], Vilmantas Giedraitis[137], Anette P Gjesing[67], Penny Gordon-Larsen[138,139], Mathias Gorski[38,66], Hans-Jörgen Grabe[140,141], Struan FA Grant[10,142,143], Niels Grarup[67], Helen L Griffiths[144], Megan L Grove[64], Vilmundur Gudnason[145,146], Stefan Gustafsson[147], Jeff Haessler[36], Hakon Hakonarson[10,142], Anke R Hammerschlag[148], Torben Hansen[67], Kathleen Mullan Harris[138,149], Tamara B Harris[150], Andrew T Hattersley[151], Christian T Have[67], Caroline Hayward[152], Liang He[153,154], Nancy L Heard-Costa[90,155], Andrew C Heath[156], Iris M Heid[38,157], Øyvind Helgeland[158,159], Jussi Hernesniemi[160,161,162], Alex W Hewitt[163,164,165], Oddgeir L Holmen[166], G Kees Hovingh[167], Joanna MM Howson[77], Yao Hu[168], Paul L Huang[24], Jennifer E Huffman[152], M Arfan Ikram[31,169,170], Erik Ingelsson[147,171], Anne U Jackson[20], Jan-Håkan Jansson[172,173], Gail P Jarvik[76,174], Gorm B Jensen[175], Yucheng Jia [14], Stefan Johansson[159,176], Marit E Jørgensen[177,178], Torben Jørgensen[57,179,180], J Wouter Jukema[181,182], Bratati Kahali[183,184,185,186], René S Kahn[187], Mika Kähönen[188,189] , Pia R Kamstrup[56], Stavroula Kanoni[22], Jaakko Kaprio[154,190,191], Maria Karaleftheri[192], Sharon LR Kardia[60], Fredrik Karpe[193,194], Sekar Kathiresan[7,24,88], Frank Kee[195], Lambertus A Kiemeney[40], Eric Kim[14], Hidetoshi Kitajima[17], Pirjo Komulainen[196,197,198], Jaspal S Kooner[83,85,199,200], Charles Kooperberg[36], Tellervo Korhonen[191,201,202], Peter Kovacs[61], Helena Kuivaniemi[81,203], Zoltán Kutalik[59,204], Kari Kuulasmaa[191], Johanna Kuusisto[205], Markku Laakso[205], Timo A Lakka[196,197,198], David Lamparter[58,59], Ethan M Lange[206], Leslie A Lange[206], Claudia Langenberg[37], Eric B Larson[76,207,208], Nanette R Lee[209,210], Terho Lehtimäki[161,162], Cora E Lewis[211], Huaixing Li[168], Jin Li[212], Ruifang Li-Gao[106], Honghuang Lin[213], Keng-Hung Lin[214], Li-An Lin[128], Xu Lin[168], Lars Lind[215], Jaana Lindström[191], Allan Linneberg[180,216,217], Ching-Ti Liu[218], Dajiang J Liu[219], Yongmei Liu[220], Ken Sin Lo[1], Artitaya Lophatananon[221], Andrew J Lotery[144], Anu Loukola[154,190], Jian'an Luan[37], Steven A Lubitz[7,24,30], Leo-Pekka Lyytikäinen[161,162], Satu Männistö[191], Gaëlle Marenne[15], Angela L Mazul[5], Mark I McCarthy[17,193,194], Roberta McKean-Cowdin[222], Sarah E

Medland[101], Karina Meidtner[223,224], Lili Milani[12], Vanisha Mistry[53,54], Paul Mitchell[225], Karen L Mohlke[206], Leena Moilanen[226], Marie Moitry[227,228], Grant W Montgomery[101,229], Dennis O Mook-Kanamori[106,230], Carmel Moore[78,231], Trevor A Mori[232], Andrew D Morris[233], Andrew P Morris[17,93], Martina Müller-Nurasyid[157,234,235], Patricia B Munroe[22,82], Mike A Nalls[236,237], Narisu Narisu[92], Christopher P Nelson[27,28], Matt Neville[193,194], Sune F Nielsen[56,57], Kjell Nikus[160], Pål R Njølstad[158,159], Børge G Nordestgaard[56,57], Dale R Nyholt[101,238], Jeffrey R O'Connel[239], Michelle L. O'Donoghue[240], Loes M Olde Loohuis[241], Roel A Ophoff[187,241], Katharine R Owen[193,194], Chris J Packard[127], Sandosh Padmanabhan[127], Colin NA Palmer[242], Nicholette D Palmer[69], Gerard Pasterkamp[110,243], Aniruddh P Patel[7,24,88], Alison Pattie[95], Oluf Pedersen[67], Peggy L Peissig[74], Gina M Peloso[218], Craig E Pennell[244], Markus Perola[191,245,246], James A Perry[239], John RB Perry[37], Tune H Pers[67,247], Thomas N Person[74], Annette Peters[224,235,248], Eva RB Petersen[249], Patricia A Peyser[60], Ailith Pirie[117], Ozren Polasek[233,250], Tinca J Polderman[148], Hannu Puolijoki[251], Olli T Raitakari[252,253], Asif Rasheed[254], Rainer Rauramaa[196,197,198], Dermot F Reilly[255], Frida Renström[129,256], Myriam Rheinberger[66], Paul M Ridker[87,88,240], John D Rioux[1,116], Manuel A Rivas[7,257], David J Roberts[78,258,259], Neil R Robertson[17,193], Antonietta Robino[80], Olov Rolandsson[172,260], Igor Rudan[233], Katherine S Ruth[261], Danish Saleheen[254,262], Veikko Salomaa[191], Nilesh J Samani[27,28], Yadav Sapkota[101], Naveed Sattar[127], Robert E Schoen[263], Pamela J Schreiner[264], Matthias B Schulze[223,224], Robert A Scott[37], Marcelo P Segura-Lepe[84], Svati H Shah[265], Wayne H-H Sheu[266,267,268], Xueling Sim[20,269], Andrew J Slater[270,271], Kerrin S Small[272], Albert Vernon Smith[145,146], Lorraine Southam[15,17], Timothy D Spector[272], Elizabeth K Speliotes[183,184,185], John M Starr[94,273], Kari Stefansson[145,274], Valgerdur Steinthorsdottir[274], Kathleen E Stirrups[22,25], Konstantin Strauch[157,275], Heather M Stringham[20], Michael Stumvoll[61,62], Liang Sun[153,154], Praveen Surendran[77], Amy J Swift[92], Hayato Tada[240,276], Katherine E Tansey[113,277], Jean-Claude Tardif[1,116], Kent D Taylor[14], Alexander Teumer[278], Deborah J Thompson[111], Gudmar Thorleifsson[274], Unnur Thorsteinsdottir[145,274], Betina H Thuesen[180], Anke Tönjes[279], Gerard Tromp[81,280], Stella Trompet[181,281], Emmanouil Tsafantakis[282], Jaakko Tuomilehto[191,283,284,285], Anne Tybjaerg-Hansen[57,133], Jonathan P Tyrer[117], Rudolf Uher[286], André G Uitterlinden[31,32], Matti Uusitupa[287], Sander W van der Laan[110], Cornelia M van Duijn[31], Nienke van Leeuwen[288,289], Jessica van Setten[47], Mauno Vanhala[201,290], Anette Varbo[56,57], Tibor V Varga[129], Rohit Varma[291], Digna R Velez Edwards[292], Sita H Vermeulen[40], Giovanni Veronesi[293], Henrik Vestergaard[67,178], Veronique Vitart[152], Thomas F Vogt[294], Uwe Völker[295,296], Dragana Vuckovic[75,135], Lynne E Wagenknecht[220], Mark Walker[297], Lars Wallentin[298], Feijie Wang[168], Carol A Wang[244], Shuai Wang[218], Yiqin Wang[168], Erin B Ware[60,299], Nicholas J Wareham[37], Helen R Warren[22,82], Dawn M Waterworth[300], Jennifer Wessel[301], Harvey D White[302], Cristen J Willer[183,184,303], James G Wilson[304], Daniel R Witte[305,306], Andrew R Wood[261], Ying Wu[206], Hanieh Yaghootkar[261], Jie Yao[14], Pang Yao[168], Laura M Yerges-Armstrong[239,307], Robin Young[77,127], Eleftheria Zeggini[15], Xiaowei Zhan[308], Weihua Zhang[83,84], Jing Hua Zhao[37], Wei Zhao[262], Wei Zhao[60], Wei Zhou[183,184], Krina T Zondervan[17,309], CHD Exome+ Consortium, EPIC-CVD Consortium, ExomeBP Consortium, Global Lipids Genetic Consortium, GoT2D Genes Consortium, InterAct, INTERVAL Study, ReproGen Consortium, T2D-Genes Consortium, The MAGIC Investigators, Understanding Society Scientific Group, Jerome I Rotter[14], John A Pospisilik[19], Fernando Rivadeneira[31,32], Ingrid B Borecki[26], Panos Deloukas[22,42], Timothy M Frayling[261], Guillaume Lettre[1,116§], Kari E North[310§], Cecilia M Lindgren[17,311§], Joel N Hirschhorn[7,9,312§], Ruth JF Loos[3,4,313§]

* These authors contributed equally to this work.
§ These authors jointly supervised this work.

AFFILIATIONS

1. Montreal Heart Institute, Universite de Montreal, Montreal, Quebec, H1T 1C8, Canada
2. Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, 37203, USA
3. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA
4. The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, 10069, USA
5. Department of Epidemiology, University of North Carolina, Chapel Hill, NC, 27514, USA
6. Human Genetics Center, The University of Texas School of Public Health, The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, TX, 77030, USA
7. Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
8. Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA
9. Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, 02115, USA
10. Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA
11. Quantinuum Research LLC, San Diego CA, 92101, USA
12. Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia
13. Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, 37203, USA
14. Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA
15. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK
16. Department of Mathematical and Statistical Sciences, University of Colorado, Denver, CO, 80204, USA
17. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
18. Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, Cyprus
19. Max Planck Institute of Immunobiology and Epigenetics, Freiburg, 79108, Germany
20. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, 48109, USA
21. McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO, 63108, USA
22. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK
23. Department of Vascular Medicine, AMC, Amsterdam, 1105 AZ, The Netherlands
24. Massachusetts General Hospital, Boston, MA, 02114, USA
25. Department of Haematology, University of Cambridge, Cambridge, CB2 0PT, UK
26. Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, 63108, USA
27. Department of Cardiovascular Sciences, Univeristy of Leicester, Glenfield Hospital, Leicester, LE3 9QP, UK
28. NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP, UK
29. Department of Medicine, Harvard University Medical School, Boston, MA, 02115, USA
30. Medical and Population Genetics Program, Broad Institute, Cambridge, MA, 02141, USA
31. Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
32. Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
33. Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA

34. Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA
35. Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA
36. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA, 98109, USA
37. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, CB2 0QQ, UK
38. Department of Genetic Epidemiology, University of Regensburg, Regensburg, D-93051, Germany
39. Netherlands Comprehensive Cancer Organisation, Utrecht, 3501 DB, The Netherlands
40. Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands
41. School of Kinesiology and Health Science, Faculty of Health, York University, Toronto
42. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia
43. Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA, 92093, USA
44. INSERM U1167, Lille, F-59019, France
45. Institut Pasteur de Lille, U1167, Lille, F-59019, France
46. Universite de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related diseases, Lille, F-59019, France
47. Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands
48. Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands
49. Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK
50. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, 53201, USA
51. INSERM U1018, Centre de recherche en Épidemiologie et Sante des Populations (CESP), Villejuif, France
52. Department of Cardiology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, 2100, Denmark
53. Metabolic Research Laboratories, University of Cambridge, Cambridge, CB2 0QQ, UK
54. NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK
55. Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, 37203, USA
56. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, 2730, Denmark
57. Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2200, Denmark
58. Department of Computational Biology, University of Lausanne, Lausanne, 1011, Switzerland
59. Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
60. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, 48109, USA
61. IFB Adiposity Diseases, University of Leipzig, Leipzig, 04103, Germany
62. University of Leipzig, Department of Medicine, Leipzig, 04103, Germany
63. Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, 14558, Germany
64. School of Public Health, Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX, 77030, USA
65. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030 USA
66. Department of Nephrology, University Hospital Regensburg, Regensburg, 93042, Germany

67. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2100, Denmark
68. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
69. Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
70. Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, 7100, Denmark
71. Institute of Regional Health Research, University of Southern Denmark, Odense, 5000, Denmark
72. MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, SE5 8AF, UK
73. NIHR Biomedical Research Centre for Mental Health, South London and Maudsley Hospital, London, BR3 3BX, UK
74. Marshfield Clinic Research Institute, Marshfield, WI, 54449, USA
75. Department of Medical Sciences, University of Trieste, Trieste, 34137, Italy
76. Department of Medicine, University of Washington, Seattle, WA, 98195, USA
77. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK
78. NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
79. American Cancer Society, Epidemiology Research Program, Atlanta, GA, 30303, USA
80. Institute for Maternal and Child Health - IRCCS “Burlo Garofolo”, Trieste, 34137, Italy
81. Weis Center for Research, Geisinger Health System, Danville, PA 17822
82. NIHR Barts Cardiovascular Research Unit, Barts and The London School of Medicine & Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK
83. Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, Middlesex, UB1 3HW, UK
84. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, W2 1PG, UK
85. Imperial College Healthcare NHS Trust, London, W12 0HS, UK
86. Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA
87. Division of Preventive Medicine, Brigham and Women's and Harvard Medical School, Boston, MA, 02215, USA
88. Harvard Medical School, Boston, MA, 02115, USA
89. Medical department, Lillebaelt Hospital, Vejle, 7100, Denmark
90. NHLBI Framingham Heart Study, Framingham, MA, 01702, USA
91. Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, 34100, Italy
92. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 20892, USA
93. Department of Biostatistics, University of Liverpool, Liverpool, L69 3GL, UK
94. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
95. Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
96. Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands
97. Department of Ophthalmology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands
98. Menzies Health Institute Queensland, Griffith University, Southport, QLD, Australia
99. Department of Biomedical Infomatics and Medical Education, University of Washington, Seattle, WA, 98195, USA
100. Diamantina Institute, University of Queensland, Brisbane, Queensland, 4072, Australia
101. QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4006, Australia

102. British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, CB2 0QQ, UK
103. Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
104. Department of Clinical Chemistry and Haematology, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, 3508 GA, The Netherlands
105. Utrecht Institute for Pharmaceutical Sciences, Division Pharmacoepidemiology & Clinical Pharmacology, Utrecht University, Utrecht, 3508 TB, The Netherlands
106. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, 2300RC, The Netherlands
107. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, 17671, Greece
108. Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, 55454, USA
109. VU University Medical Center, Department of Internal Medicine, Amsterdam, 1007 MB, The Netherlands
110. Laboratory of Experimental Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
111. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK
112. Institute of Cardiovascular Science, University College London, London, WC1E 6JF, UK
113. MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, BS8 2BN, UK
114. Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA, 98109, USA
115. Memorial Sloan Kettering Cancer Center, Department of Epidemiology and Biostatistics, New York, NY, 10017, USA
116. Department of Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, H3T 1J4, Canada
117. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK
118. Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
119. Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, W2 1PG, UK
120. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, 45110, Greece
121. Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, UK
122. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104, USA
123. Department of Ophthalmology, University of Cologne, Cologne, 50937, Germany
124. CNR Institute of Clinical Physiology, Pisa, Italy
125. Department of Clinical & Experimental Medicine, University of Pisa, Italy
126. Toulouse University School of Medicine, Toulouse, TSA 50032 31059, France
127. University of Glasgow, Glasgow, G12 8QQ, UK
128. Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, 77030, USA
129. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmo, SE-20502, Sweden
130. Department of Nutrition, Harvard School of Public Health, Boston, MA, 02115, USA
131. Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, 901 87, Sweden

132. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, 17475, Germany
133. Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, 2100, Denmark
134. Ilaria Gandin, Research Unit, AREA Science Park, Trieste, 34149, Italy
135. Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy
136. Centre for Biological Sciences, Faculty of Natural and Environmental Sciences, University of Southampton, Southampton, SO17 1BJ, UK
137. Geriatrics, Department of Public Health, Uppsala University, Uppsala, 751 85, Sweden
138. Carolina Population Center, University of North Carolina, Chapel Hill, NC, 27514, USA
139. Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 27514, USA
140. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, 17475, Germany
141. German Center for Neurodegenerative Diseases (DZNE), Rostock/Greifswald, Greifswald, 17475, Germany
142. Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA
143. Division of Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA
144. Vision Sciences, Clinical Neurosciences Research Group, Faculty of Medicine, University of Southampton, Southampton, SO16 6YD, UK
145. Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland
146. Icelandic Heart Association, Kopavogur, 201, Iceland
147. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, 751 41, Sweden
148. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, 1081 HV, The Netherlands
149. Department of Sociology, University of North Carolina, Chapel Hill, NC, 27514, USA
150. Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, MD, 20892, USA
151. University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
152. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
153. Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC, 27708, USA
154. Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland
155. Department of Neurology, Boston University School of Medicine, Boston, MA, 02118, USA
156. Department of Psychiatry, Washington University, Saint Louis, MO, 63110, USA
157. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, 85764, Germany
158. Department of Pediatrics, Haukeland University Hospital, Bergen, 5021, Norway
159. KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, 5020, Norway
160. Department of Cardiology, Heart Center, Tampere University Hospital, and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, 33521, Finland
161. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, 33521, Finland
162. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere 33014, Finland
163. Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, Victoria, 3002, Australia

164. Centre for Ophthalmology and Vision Science, Lions Eye Institute, University of Western Australia, Perth, Western Australia, 6009, Australia
165. Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, 7000, Australia
166. K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, 7600, Norway
167. AMC, Department of Vascular Medicine, Amsterdam, 1105 AZ, The Netherlands
168. Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Shanghai, People's Republic of China, Shanghai, 200031, China
169. Department of Neurology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
170. Department of Radiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
171. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, 943 05, USA
172. Department of Public Health & Clinical Medicine, Umeå University, Umeå, SE-90185, Sweden
173. Research Unit Skellefteå, Skellefteå, SE-93141, Sweden
174. Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA
175. The Copenhagen City Heart Study, Frederiksberg Hospital, Frederiksberg, 2000, Denmark
176. Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, 5021, Norway
177. National Institute of Public Health, University of Southern Denmark, Copenhagen, 1353, Denmark
178. Steno Diabetes Center Copenhagen, Gentofte, 2800, Denmark
179. Faculty of medicine, Aalborg University, Aalborg, DK-9000, Denmark
180. Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, DK-2600, Denmark
181. Department of Cardiology, Leiden University Medical Center, Leiden, 2333, The Netherlands
182. The Interuniversity Cardiology Institute of the Netherlands, Utrecht, 2333, The Netherlands
183. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, 48109, USA
184. Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109, USA
185. Division of Gastroenterology, University of Michigan, Ann Arbor, MI, 48109, USA
186. Centre for Brain Research, Indian Institute of Science, Bangalore 560012, India
187. Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, 3584 CG, The Netherlands
188. Department of Clinical Physiology, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, 33014, Finland
189. Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland
190. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, FI-00014, Finland
191. National Institute for Health and Welfare, Helsinki, FI-00271, Finland
192. Echinos Medical Centre, Echinos, Greece
193. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 7LE, UK
194. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK
195. UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast, UK, BT12 6BJ, UK
196. Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, 70100, Finland
197. Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus, 70210, Finland

198. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland
199. National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus, London, W12 0NN, UK
200. MRC-PHE Centre for Environment and Health, Imperial College London, London, W2 1PG, UK
201. University of Eastern Finland, Kuopio, 70210, Finland
202. University of Helsinki, Helsinki, 00100, Finland
203. Department of Psychiatry, and Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Western Cape, 7505, South Africa
204. Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, 1010, Switzerland
205. Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, 70210, Finland
206. Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599, USA
207. Kaiser Permanente Washington Health Research Institute Seattle WA 98101
208. Department of Health Services, University of Washington, Seattle WA 98101
209. Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, 6000, Philippines
210. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, 6000, Philippines
211. Division of Preventive Medicine University of Alabama at Birmingham, Birmingham, AL 35205, USA
212. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Palo Alto, CA, 94304, USA
213. Department of Medicine, Boston University School of Medicine, Boston, MA, 02118, USA
214. Department of Ophthalmology, Taichung Veterans General Hospital, Taichung, Taiwan 407, Taiwan
215. Uppsala University, Uppsala, 75185, Sweden
216. Department of Clinical Experimental Research, Rigshospitalet, Copenhagen, DK-2200, Denmark
217. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2200, Denmark
218. Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA
219. Department of Public Health Sciences, Institute for Personalized Medicine, the Pennsylvania State University College of Medicine, Hershey, PA, 17033, USA
220. Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA
221. Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, CV4 7AL, UK
222. Department of Preventive Medicine, Keck School of Medicine of the University of California, Los Angeles, CA, 90089, USA
223. Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Nuthetal, 14558, Germany
224. German Center for Diabetes Research, München-Neuherberg, 85764, Germany

225. Westmead Millennium Institute of Medical Research, Centre for Vision Research and Department of Ophthalmology, University of Sydney, Sydney, New South Wales, 2022, Australia
226. Department of Medicine, Kuopio University Hospital, Kuopio, 70210, Finland
227. Department of Epidemiology and Public Health, University of Strasbourg, Strasbourg, F-67085, France
228. Department of Public Health, University Hospital of Strasbourg, Strasbourg, F-67081, France

229. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, 4072, Australia
230. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, 2300RC, The Netherlands
231. INTERVAL Coordinating Centre, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK
232. School of Medicine and Pharmacology, The University of Western Australia, Perth, Western Australia, 6009, Australia
233. Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK
234. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universitat, Munich, 81377, Germany
235. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, 80802, Germany
236. Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD, 20892, USA
237. data tecnica international, Glen Echo, MD, USA
238. Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, 4059, Australia
239. Program for Personalized and Genomic Medicine, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, 21201, US
240. Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA
241. Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA, 90095, USA
242. Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Medical Research Institute, Ninewells Hospital and Medical School, Dundee, DD1 9SY, UK
243. Laboratory of Clinical Chemistry and Hematology, Division Laboratories and Pharmacy, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
244. School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, 6009, Australia
245. University of Helsinki, Institute for Molecular Medicine (FIMM) and Diabetes and Obesity Research Program, Helsinki, FI00014, Finland
246. University of Tartu, Estonian Genome Center, Tartu, Estonia, Tartu, 51010, Estonia
247. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, 2200, Denmark
248. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, 85764, Germany
249. Department of Clinical Immunology and Biochemistry, Lillebaelt Hospital, Vejle, 7100, Denmark
250. School of Medicine, University of Split, Split, 21000, Croatia
251. South Ostrobothnia Central Hospital, Seinajoki, 60220, Finland
252. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, 20521, Finland
253. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, 20520, Finland
254. Centre for Non-Communicable Diseases, Karachi, Pakistan
255. Merck, Sharp & Dohme, Genetics and Pharmacogenomics, Boston, MA, 02115, USA
256. Department of Biobank Research, Umeå University, Umeå, SE-90187, Sweden
257. Nuffield Department of Clinical Medicine, Oxford, OX37 BN, UK
258. NHS Blood and Transplant - Oxford Centre, Oxford, OX3 9BQ, UK
259. BRC Haematology Theme and Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 9DU, UK
260. Department of Public Health and Clinical Medicine, Unit of Family Medicine, Umeå University, Umeå, 90185, Sweden

261. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK
262. Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA
263. University of Pittsburgh Medical Center, Departments of Medicine and Epidemiology, Pittsburgh, PA, 15213, USA
264. Division of Epidemiology & Community Health University of Minnesota, Minneapolis, MN, 55454, USA
265. Duke Molecular Physiology Institute, Duke University, Durham NC, 27701
266. Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan 407, Taiwan
267. School of Medicine, National Defense Medical Center, Taipei, Taiwan 114, Taiwan
268. School of Medicine, National Yang-Ming University, Taipei, Taiwan
269. Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore 117549, Singapore
270. Genetics, Target Sciences, GlaxoSmithKline, Research Triangle Park, NC, 27709, US
271. OmicSoft a QIAGEN Company, Cary, NC, 27513, US
272. Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK
273. Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, EH8 9JZ, UK
274. deCODE Genetics/Amgen inc., Reykjavik, 101, Iceland
275. Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, 81377, Germany
276. Kanazawa University, Kanazawa, 920-8641, Japan
277. College of Biomedical and Life Sciences, Cardiff University, Cardiff, CF14 4EP, UK
278. Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany
279. Center for Pediatric Research, Department for Women's and Child Health, University of Leipzig, Leipzig, 04103, Germany
280. Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Western Cape, 7505, South Africa
281. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, 2333, The Netherlands
282. Anogia Medical Centre, Anogia, Greece
283. Centre for Vascular Prevention, Danube-University Krems, Krems, 3500, Austria
284. Dasman Diabetes Institute, Dasman, 15462, Kuwait
285. Diabetes Research Group, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
286. Department of Psychiatry, Dalhousie University, Halifax, B3H 4R2, Canada
287. Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, 70210, Finland
288. Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, 1081BT, The Netherlands
289. Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, 2333ZC, The Netherlands
290. Central Finland Central Hospital, Jyvaskyla, 40620, Finland
291. USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, CA, 90033, USA
292. Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, 37203, USA
293. Research Center on Epidemiology and Preventive Medicine, Department of Medicine and Surgery, University of Insubria, Varese, 21100, Italy
294. Merck, Sharp & Dohme, Cardiometabolic Disease, Kenilworth, NJ, 07033, USA

295. DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, 17475, Germany
296. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, 17475, Germany
297. Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, NE2 4HH, UK
298. Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala University, Uppsala, 752 37, Sweden
299. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104, USA
300. Genetics, Target Sciences, GlaxoSmithKline, King of Prussia, PA, US
301. Departments of Epidemiology & Medicine, Diabetes Translational Research Center, Fairbanks School of Public Health & School of Medicine, Indiana University, Indiana, IN, 46202, USA
302. Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand
303. Department of Human Genetics, University of Michigan, Ann Arbor, MI, 48109, USA
304. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, 39216, USA
305. Danish Diabetes Academy, Odense, 5000, Denmark
306. Department of Public Health, Aarhus University, Aarhus, 8000, Denmark
307. GlaxoSmithKline, King of Prussia, PA, 19406, USA
308. Department of Clinical Sciences, Quantitative Biomedical Research Center, Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA
309. Endometriosis CaRe Centre, Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford, OX3 9DU, UK
310. Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, NC, 27514, USA
311. Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford, OX3 7BN, UK
312. Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, 02115, USA
313. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10069, USA

Correspondence to:

Ruth J. F. Loos (ruth.loos@mssm.edu)

Joel N. Hirschhorn (joelh@broadinstitute.org)

ABSTRACT

Genome-wide association studies (GWAS) have identified >250 loci for body mass index (BMI), implicating pathways related to neuronal biology. Most GWAS loci represent clusters of common, non-coding variants from which pinpointing causal genes remains challenging. Here, we combined data from 718,734 individuals to discover rare and low-frequency (MAF<5%) coding variants associated with BMI. We identified 14 coding variants in 13 genes, of which eight in genes (*ZBTB7B, ACHE, RAPGEF3, RAB21, ZFHX3, ENTPD6, ZFR2, ZNF169*) newly implicated in human obesity, two (*MC4R, KSR2*) previously observed in extreme obesity, and two variants in *GIPR*. Effect sizes of rare variants are ~10 times larger than of common variants, with the largest effect observed in carriers of an *MC4R* stop-codon (p.Tyr35Ter, MAF=0.01%), weighing ~7kg more than non-carriers. Pathway analyses confirmed enrichment of neuronal genes and provide new evidence for adipocyte and energy expenditure biology, widening the potential of genetically-supported therapeutic targets to treat obesity.

Obesity is a heritable disease and represents a major unmet public health problem with only a few safe and long-term effective therapies[1] and intervention strategies[2]. To understand the genetic basis of obesity and identify potential targets for new therapies, genome-wide association studies (GWAS) for body mass index (BMI) and obesity risk have identified >250 common variants over the past decade[3-7]. Consistent with single-gene disorders of obesity[8], tissue expression and gene-set enrichment analyses for genes in BMI-associated loci have shown that the central nervous system (CNS) plays a critical role in body weight regulation[5]. While the numerous GWAS loci have provided insight into broad biological mechanisms underlying body weight regulation, pinpointing the causal gene(s)/variant(s) remains a major challenge[9], as GWAS-identified variants are typically non-coding and may affect genes at long distance. The association of intronic *FTO* variants with BMI illustrates the challenges of identifying causal regulatory effects. The proposed causal variant was found to regulate the expression of nearby *RPGRIP1L* in some studies[10-12], whereas others found that it regulates distant *IRX3/IRX5* genes in specific cell types[13,14].

To expedite mapping of obesity-related genes, we performed an exome-wide search for low-frequency (LF, minor allele frequency [MAF]=1-5%) and rare (R, MAF<1%) single nucleotide variants (SNVs) associated with BMI using exome-targeted genotyping arrays. A total of 125 studies ($N_{individuals}$=718,734) performed single-variant association between up to 246,328 SNVs and BMI. In addition, we performed gene-based meta-analyses to aggregate rare and LF (R/LF) coding SNVs across 14,541 genes. Using genetic, functional and computational follow-up analyses, we gained insights into the function of BMI-implicated genes, and the biological pathways through which they may influence body weight.

RESULTS

Fourteen rare and low-frequency coding variants in 13 genes

Our study comprises a discovery and a follow-up stage (**Supplementary Figure 1**, **Supplementary Tables 1-3, Online Methods**). In our primary analysis, the discovery stage includes data from 123

studies (N_{max}=526,508) across five ancestry groups, predominantly European (~85%). Each study performed single-variant association analyses of coding variants present on the exome array, including up to 13,786 common (MAF>5%) and 215,917 R/LF coding SNVs (exons and splicing sites). Summary statistics were combined using fixed-effect meta-analyses. SNV-associations of R/LF variants that reached suggestive significance ($P<2.0x10^{-6}$) were taken forward for follow-up in two European cohorts, deCODE (N_{max}=72,613) and UK Biobank (N_{max}=119,613 [interim release]). Overall significance was assessed after combining results of discovery and follow-up studies into a final meta-analysis (all-ancestries, sex-combined, additive model, N_{max}=718,734), SNV-associations that reached $P<2x10^{-7}$ were considered array-wide significant[15,16] (**Table 1**, **Supplementary Table 4, Supplementary Figures 2-4**). In secondary analyses, we performed sex-specific analyses, analyses limited to individuals of European ancestry, and analyses using a recessive model.

In our primary analysis of R/LF variants, we identified five rare SNVs in three genes (*KSR2*, 2 in *MC4R*, 2 in *GIPR*) and nine LF SNVs in eight genes (*ZBTB7B, 2 in ACHE, RAPGEF3, PRKAG1, RAB21, HIP1R, ZFHX3*, *ENTPD6*) (**Table 1**, **Box 1**, **Supplementary Table 5, Supplementary Figure 3a**). In secondary analyses, we identified two additional LF SNVs, one in all-ancestry women-only (*ZFR2*) and one in European ancestry only analyses (*ZNF169*) (**Table 1**, **Supplementary Tables 6-8, Supplementary Figures 3b, 3c**). Of these 16 SNVs, located in 13 genes, the two SNVs in *MC4R* (r^2=1, *D'*=1) and two in *ACHE* (r^2=0.98, *D'*=0.99) were in high LD, whereas the two SNVs in *GIPR* (r^2=0, *D'*=0.16) were independent of each other. Hence, the 16 SNVs represent 14 independent SNVs (4 rare, 10 LF), of which eight locate in genes not previously implicated in BMI (*ZBTB7B, ACHE, RAPGEF3, RAB21, ZFHX3, ENTPD6, ZFR2, ZNF169*), and six are located in five loci that were previously identified by GWAS (*PRKAG1/BCDIN3D, HIP1R/CLIP1, MC4R, GIPR/QPCTL*)[5] and/or through sequencing of severe early-onset obesity cases (*MC4R, KSR2*)[17-19] (**Supplementary Figure 5**). Conditional analyses established that coding SNVs in *PRKAG1, MC4R* and *GIPR* are independent of the common lead variants in GWAS loci (rs7138803, rs17782313, rs2287019, respectively), whereas the SNV in *HIP1R* and GWAS locus near *CLIP1* (rs11057405) represent the same signal (**Online Methods, Supplementary**

Tables 9, 10, Supplementary Figure 5).

Next, we performed gene-based association tests (SKAT, VT, broad, strict) in up to 14,541 genes[20] to examine whether these aggregated analyses would yield new evidence for multiple R/LF coding SNVs in the same gene affecting BMI (**Online Methods**). Using broad SNV inclusion criteria, associations for 13 genes reached array-wide significance ($P<2.5x10^{-6}$)[15,16], four of which had not been highlighted in single-variant analyses (**Table 2, Supplementary Table 11**). Conditional analyses showed that only for *GIPR* was the gene-based association driven by multiple SNVs (**Table 2**, **Supplementary Table 12**). For all other genes, associations were driven by a single SNV only, but these SNVs had not reached array-wide significance in single-variant analyses.

Taken together, we identified 14 R/LF coding SNVs in 13 genes that are independently associated with BMI, four rare SNVs in three genes, and 10 LF SNVs in 10 genes. One SNV (*ZFR2*) showed a sex-specific effect, whereas no ancestry-specific effects were observed (**Supplementary Note**, **Supplementary Tables 6-8**, **Supplementary Figure 6**). Eight (*ACHE, ENTPD6, RAB21, RAPGEF3, ZBTB7B, ZFHX3, ZFR2, ZNF169*) of these 13 genes have not been previously implicated in body weight regulation (**Table 1**).

Novel common coding variants associated with BMI

Although the main focus of our study was on R/LF coding SNVs, we also identified 92 common coding variants ($P<2.0x10^{-7}$, **Supplementary Tables 4, Supplementary Figures 4, 7**), of which 41 were novel (**Supplementary Table 9, Supplementary Note**). These novel common loci had not been identified in previous GWAS efforts, because our current sample size is more than twice as large as the most recent GWAS meta-analysis[5], and also because some SNVs were not tested before, as they were not present on the HapMap reference panel and/or were on the X-chromosome, which was not analyzed. Because of the increased samples size, effect sizes of the 41 novel common loci are smaller (on average 0.014 SD/allele, [range: 0.010–0.024]) than of previously established common loci (0.021 SD/allele, [0.010–0.050]) (**Supplementary Figure 7**).

Impact of R/LF SNVs on BMI and obesity risk

The minor allele for half of the 14 R/LF SNVs is associated with lower BMI (**Table 1, Figure 1**). The effects of LF SNVs range between 0.024 and 0.066 SD/allele, equivalent to ~0.11 to 0.30 kg/m^2 in BMI or ~0.315 to 0.864 kg in body weight for a 1.7m tall person. Effects of rare SNVs range between 0.06 and 0.54 SD per allele, equivalent to 0.26 to 2.44 kg/m^2 or 0.74 kg to 7.05 kg per allele (**Table 1, Figure 1**). By comparison, these rare SNV effect sizes are on average ten times larger than those for previously identified GWAS loci (effect$_{mean}$=0.019 SD/allele, ~0.086 kg/m^2 or ~0.247 kg/allele) of which the largest effect is seen for the *FTO* locus (0.08 SD/allele, ~0.35 kg/m^2 or 1 kg/allele) and those for other GWAS loci range between 0.010 and 0.056 SD/allele (~0.045 to 0.25 kg/m^2, or 0.130 to 0.728 kg)[5].

Effect sizes increase as MAF decreases, in particular for SNVs with a MAF<0.5% (~1 heterozygote carrier in 100 people), consistent with the statistical power of our sample (**Figure 1**). For example, the nonsense p.Tyr35Ter *MC4R* SNV (rs13447324, MAF=0.01%) is present in ~1 in 5,000 individuals and results in a ~7 kg higher body weight for a 1.7m tall person. The two *GIPR* SNVs contribute independently to a *lower* body weight, carriers (1 in ~455 individuals) of p.Arg190Gln (rs139215588) weigh ~1.92 kg (0.148 SD BMI) less than non-carriers and carriers (1 in ~385 individuals) of p.Glu288Gly (rs143430880) weigh ~1.99 kg (0.153 SD BMI) less. Among 115,611 individuals of the UK Biobank, one apparently healthy 61-year-old woman, with no reported illnesses, carried both rare *GIPR* alleles and weighed ~11.2 kg less (equivalent to -0.86 SD BMI or 3.87 kg/m^2) than the average non-carrier of the same height (**Supplementary Figure 8)**. The possible synergistic effect of the two *GIPR* alleles needs confirmation by additional individuals that carry both variants.

Even though effect sizes of LF and, in particular, rare SNVs tend to be larger than those of common GWAS-identified loci[5], the 14 SNVs combined explain <0.1% of BMI variation, because of their low population frequency (**Table 1**, **Online Methods**). Also, although the effects of the four rare SNVs (*KSR2*, *MC4R*, 2 in *GIPR*) are large by GWAS standards, penetrance for obesity is still expected to be low. Indeed, using data from the UK Biobank (N$_{max}$=119,781), we compared the prevalence of normal-weight (18.5 kg/m^2 ≤ BMI < 25 kg/m^2) and obesity (BMI ≥ 30 kg/m^2) between carriers and non-carriers

(**Supplementary Table 13**, **Online Methods**). For *GIPR* (p.Arg190Gln, p.Glu288Gly), both BMI-decreasing SNVs, carriers tended (P<0.05) to have a lower obesity prevalence (21.2%, 20.1%, respectively), compared to non-carriers (25.1%, 25%). For *MC4R* p.Tyr35Ter and *KSR2* p.Arg525Gln, the prevalence of obesity between carriers (30%, 25.7%, resp.) and non-carriers (25.1%, 25.3%) was not significantly different.

We examined whether R/LF SNVs affect obesity risk early on in life by combining data from three case-controls studies of childhood obesity (N_{cases}=4,395, $N_{controls}$=13,072) (**Online Methods, Supplementary Table 14**). Associations for 10 of 13 SNVs were directionally consistent with those observed for BMI in adults (77%, $P_{binomial}$=0.046), three of which (*ZBTB7B, PRKAG1, RAB21*) reached nominal significance (P<0.05). While no carriers of the *MC4R* mutations were available for analyses, the role of *MC4R* in body weight regulation in childhood was established almost two decades ago[17,19,21].

Impact of R/LF SNVs on cardiometabolic and other traits

To examine whether identified SNVs affect other traits, we obtained results from multiple large-scale genetic consortia (GIANT[15], MAGIC, GoT2D/T2D-GENES[16], GLGC, ICBP[22], REPROGEN[23]) (**Supplementary Table 15, Supplementary Figure 9)**, and performed phenome-wide association (PheWAS) analyses using electronic medical record (EMR) data from BioVu and UK Biobank (**Online Methods**, **Supplementary Table 16**). The BMI-increasing allele of *ZBTB7B* p.Pro190Ser is associated with greater height, and those of *PRKAG1, ACHE*, and *RAPGEF3* SNVs are associated with shorter height, but association with other traits differ. Specifically, *PRKAG1* p.Thr38Ser Ser-allele carriers appear heavier and shorter, have lower HDL-cholesterol levels, earlier age at menarche (reported before[23]) and higher systolic blood pressure, which is in agreement with PheWAS analyses showing an increased risk of "malignant essential hypertension" and "hypertension" (**Supplementary Table 16**). While carriers of the *RAPGEF3* p.Leu300Pro Pro-allele are also heavier and shorter, they have a lower WHR_{adjBMI}[24] and lower fasting insulin levels (**Supplementary Table 15),** consistent with PheWAS results that show lower odds of "secondary diabetes mellitus" (**Supplementary Table 16**). Thus, while all SNVs are associated

with BMI, their patterns of association with other traits suggest they may affect different physiological pathways.

Gene set enrichment analyses

To test whether the R/LF variants implicate biological pathways, we performed gene set enrichment analyses. Similar to our previous analysis of GWAS for BMI[5], we analyzed coding variants that reached $P<5x10^{-4}$, using a DEPICT version adapted for exome-array analysis[15] (**Online Methods, Supplementary Note**). We used 50 R/LF coding variants as input (all $P<5x10^{-4}$, **Online Methods**) and observed significant enrichment (**Figure 2, Supplementary Table 17, Supplementary Figure 10a**). Many of these relate to neuronal processes, such as neurotransmitter release and synaptic function (e.g. glutamate receptor activity, regulation of neurotransmitter levels, synapse part), consistent with previous findings from GWAS[5]. When we excluded variants near (+/- 1Mb) previously identified GWAS loci, we still observed 29 significantly enriched gene sets (in 12 meta-gene sets) (**Supplementary Table 18, Supplementary Figure 10b**), thereby providing an independent confirmation of the GWAS gene set enrichment results. In addition to neuronal-related gene sets, the analyses with R/LF coding variants newly identified a cluster of metabolic pathways related to insulin action and adipocyte/lipid metabolism (e.g. enhanced lipolysis, abnormal lipid homeostasis, increased circulating insulin level, **Figure 2**). Finally, we observed that R/LF BMI-associated coding variants are more effective at identifying enriched gene sets compared to common coding variants. Specifically, adding 192 common coding SNVs (all $P<5x10^{-4}$) to the analysis decreased the number of enriched gene sets from 471 (106 meta-gene sets) seen with R/LF coding SNVs to 62 (24 meta-gene sets) (**Supplementary Table 19, Supplementary Figure 10c**). We observed fewer significant genes sets with the combined common and R/LF analysis, despite including more total coding variants and a higher fraction of array-wide significant coding variants. One possible explanation is that R/LF coding variants may fall in the causal gene more often than do common coding variants, which suggests that the R/LF variants are more likely to be causal, rather than simply in LD with causal variants.

We also used gene set enrichment analysis to prioritize candidate genes. Among the genes with R/LF coding variants associated with BMI at $P<5x10^{-4}$, a subset is prominently represented in the CNS-related enriched gene sets (**Figure 2**) and is proposed to influence neurotransmission and/or synaptic organization, function and plasticity. These include genes in regions with suggestive evidence of association from GWAS (e.g. *CARTPT, MAP1A, ERC2*) and genes in regions not previously implicated by GWAS (e.g. *CALY, ACHE, PTPRD, GRIN2A*). The non-neuronal metabolic gene sets implicate two genes (*CIDEA, ADH1B*) that are markers of brown or "beige" adipose tissue[25,26], providing new supporting evidence for a causal role of this aspect of adipocyte biology.

***Drosophila* fly results**

To test for potential adiposity-driving effects of gene regulation, we performed tissue-specific RNAi-knockdown experiments in *Drosophila*. We generated adipose-tissue (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdown crosses for nine of the 13 candidate genes for which fly orthologues exist (**Supplementary Table 20**) and performed whole body triglyceride analysis in young adult male flies. Triglycerides, the major lipid storage form in animals, were chosen as a direct measure of fly adiposity. Both neuronal and fat-body knockdown of *zfh2*, the orthologue of *ZFHX3*, resulted in significantly increased triglyceride levels. Adipose-tissue specific, but not neuronal, knockdown of *epac* (*RAPGEF3*) was lethal. Tissue-specific loss-of-function of the other seven genes tested did not affect triglyceride levels.

R/LF coding SNVs in monogenic and syndromic genes

We identified 39 genes in the literature that have been convincingly implicated in monogenic obesity or syndromes of which obesity is one of the main features (**Supplementary Table 21, 22, Supplementary Figure 11**). Of the 652 R/LF SNVs in these 39 monogenic and/or syndromic genes, five R/LF SNVs were significantly associated with BMI (Bonferroni-corrected *P*-value = $7.7x10^{-5}$ (=0.05/652)). Beside SNVs in *MC4R* (p.Tyr35Ter, Asp37Val) and *KSR2* (Arg525Gln), already highlighted in the single-variant analyses, we identified an additional SNV in *MC4R* (p.Ile251Leu) and one in *BDNF* (p.Glu6Lys). *MC4R*

p.Ile251Leu has been previously shown to protect against obesity[27], whereas *BDNF* p.Glu6Lys, independent of previously GWAS-identified SNVs (r^2=0.01, D'=1.0)[5], has not been implicated in body weight regulation before. We examined whether the 652 R/LF SNVs showed enrichment for association with BMI compared to R/LF coding SNVs in all other genes, but found no evidence to support this.

DISCUSSION

In this meta-analysis of exome-targeted genotyping data, we identified 14 R/LF coding variants in 13 genes associated with BMI. Eight of these genes (*ACHE, ENTPD6, RAB21, RAPGEF3, ZBTB7B, ZFHX3, ZFR2, ZNF169*) have not been previously implicated in human obesity, but evidence from animal studies provides support for a role in energy metabolism for some of these, such as *ACHE*[28,29], *RAPGEF3*[30-33], and *PRKAG1*[34-39]. Others fall into established BMI GWAS loci (*PRKAG1/BCDIN3D, HIP1R/CLIP1, MC4R, GIPR/QPCTL*)[5] and/or were previously implicated in severe early-onset obesity (*MC4R, KSR2*)[17-19] and using this exome-targeted approach, we pinpoint R/LF variants in these loci that play a role in obesity in the general population. Pathway analyses confirm a key role for neuronal processes, and newly implicate adipocyte and energy expenditure biology.

Consistent with other polygenic traits[15,23,40-43], we show that large sample sizes are needed to identify R/LF variants. Observed effect sizes reflect the statistical power of our sample size, and are particularly large for SNVs with a MAF < 0.05%. The existence of rare alleles with larger effects on BMI than have been observed for common alleles might reflect negative or stabilizing selection on the extremes of BMI. However, rare variants with smaller effects almost certainly exist, larger samples will be needed to uncover these. Our study was limited to coding variants on the exome-array, large-scale sequencing studies will be needed to test for variants not covered by exome-arrays.

The strongest association was observed for a stop-codon (p.Tyr35Ter, rs13447324, MAF= 0.01%) in *MC4R*, with carriers weighing on average 7kg more than non-carriers. *MC4R* is widely expressed in the CNS and is an established key player in energy balance regulation[44,45]. Mouse and human

studies showed already two decades ago that MC4R-deficiency results in extreme obesity, mainly through increased food intake[46-49]. p.Tyr35Ter, which results in MC4R-deficiency[51], was one of the first *MC4R* mutations discovered in monogenic cases of obesity[17,19], in whom the mutation is >20x more prevalent than in the general population[17,50,52,53]. Here, we show that p.Tyr35Ter plays a role outside the setting of early-onset and extreme obesity. Despite its large effect, penetrance is low, and does not fit the model of a fully penetrant Mendelian variant.

While significant R/LF coding variants are strong candidates for being causal, the strongest implication of causal genes is provided by association with multiple independent coding variants, as we demonstrate for *GIPR*. We identified two rare variants in *GIPR* (p.Arg190Gln, rs139215588, MAF=0.11%, p.Glu288Gly, rs143430880, MAF=0.13%) independently associated with lower BMI, carriers of either variant weigh ~2 kg less than non-carriers. Common variants in/near *GIPR* have been found to associate with lower BMI[55] and delayed glucose and insulin response to an oral glucose challenge[54]. However, the two rare variants influence BMI independently of these common ones and are not associated with type 2 diabetes or glycemic traits tested. Rodent models have provided strong evidence for a role of GIPR in body weight regulation. *Gipr*-deficient mice are protected from diet-induced obesity[56] and have an increased resting metabolic rate[57]. Blocking GIP-signaling using a vaccination approach in mice on a high-fat diet reduces weight gain, mainly through reduced fat accumulation, mediated through increased energy expenditure[58]. Manipulation of incretins (GIP, GLP1) and their receptors has complex effects on obesity and insulin secretion/action that may differ between human and mice[59]. The human genetic data suggest that inhibition of GIPR-signaling might present a therapeutic target for the treatment of obesity[60].

A fourth rare variant, in *KSR2*, (p.Arg525Gln, rs56214831, MAF=0.82%) increases body weight by ~740g/allele. *KSR2* is another gene previously implicated in energy metabolism and obesity[18,61,62]. In a recent study, mutation carriers were hyperphagic, had a reduced basal metabolic rate and severe insulin resistance[18]. Consistent with human data, *Ksr2*$^{-/-}$ mice were obese, hyperphagic, and had a reduced energy expenditure[18,61-63]. KSR2 is almost exclusively expressed in the brain and interacts with multiple

proteins[64], including AMP-activated protein kinase (AMPK), a key regulator of energy homeostasis[61,62]. Interestingly, *KSR2* is one of the first genes implicated in severe, early-onset obesity in which mutations not only affect food intake but also basal metabolic rate, and is thought to act via neuronal effects[18] (**Figure 2**).

Despite convincing associations of these four rare variants in *MC4R*, *GIPR* and *KSR2*, their penetrance for obesity is low (**Supplementary Table 13**). This is consistent with the polygenic and multifactorial nature of obesity, where variants across a range of frequencies and effect sizes contribute to the phenotype in any one person. Despite low predictive power, it remains possible that the identities of particular variants in any one person may contribute to different balances of underlying physiologies and hence, different responses to treatments. This was illustrated in two patients with monogenic obesity due to *POMC* mutations, these patients lack the main activator of MC4R and were effectively treated with an MC4R-agonist[65].

Of the coding variants in newly identified genes, some have well-known connections to obesity. For example, *PRKAG1* encodes the γ1-subunit of AMPK, a critical cellular energy sensor[34]. In the hypothalamus, AMPK integrates hormonal and nutritional signals with neuronal networks to regulate food intake and whole-body energy metabolism[35-37]. Furthermore, hypothalamic AMPK is a key regulator of brown adipose tissue in mice[36,38,39]. The BMI-decreasing allele at the associated *PRKAG1* variant (p.Thr38Ser, rs1126930, MAF=3.22%) has additional beneficial effects on blood pressure, providing additional genetic support for modulation of AMPK as an ongoing therapeutic avenue for treatment.

ACHE, in which p.His353Asn (rs1799805, MAF=3.9%) is associated with increased BMI, is another candidate gene related to neuronal biology, involved in the signaling of acetylcholine at neuromuscular junction and brain cholinergic synapses[67,68]. Inhibitors of ACHE, used to treat moderate-to-severe Alzheimer's Disease[69], results in weight loss in humans and Ache-deficient mice have delayed weight gain[28,29]. However, these may be indirect consequences of adverse gastrointestinal and neuromuscular effects, respectively[28,29,70,71].

Another LF coding variant (p.Leu300Pro, rs145878042, MAF=1.1%) is located in *RAPGEF3*,

and has strong effects on multiple other phenotypes. The BMI-increasing 300Pro-allele is associated with shorter height, lower WHR_{adjBMI} and lower insulin levels, suggesting that this variant has multiple physiologic consequences. Data from animal models also suggest complex effects of RAPGEF3 on adipocyte biology, energy balance and glucose metabolism[30-33]. For example, in one study, global deletion of *Rapgef3* in mice on a high-fat diet are resistant to obesity due to reduced food intake and have an increased glucose tolerance[31]. However, in a similar study, *Rapgef3*$^{-/-}$ mice develop severe obesity, increased respiratory exchange ratio and impaired glucose tolerance[33]. Adipose tissue-specific *Rapgef3* knockout mice on a high-fat diet are also more prone to obesity, show increased food intake, reduced energy expenditure, impaired glucose tolerance, and reduced circulating leptin levels[72]. More research is needed to understand the consequences of *RAPGEF3* manipulation.

The remaining genes with significant associations, *ENTPD6*, *HIP1R*, *RAB21*, *ZFR2*, *ZBTB7*, and *ZFHX3*, have no clear prior evidence for a role in energy homeostasis, and in-depth functional follow up is needed to gain insight in how they affect body weight. Here, we performed gene set enrichment analyses to better understand the biology implicated by our genetic data, and confirm the importance of neuronal processes, in particular synaptic function and neurotransmitter release, providing an independent validation of previous GWAS findings[5]. The combination of gene set enrichment and association analyses of coding variants also enables us to highlight candidate genes that are both within these gene sets and show association with BMI at R/LF coding variants. These include genes reaching array-wide significance (e.g. *ACHE, ZFR2*), and others with clear prior evidence for a role in body weight regulation (e.g. *CARTPT*[73]), but that had not been highlighted in our single-variant or gene-based association analyses. Of note, the enrichment signals were stronger with R/LF coding variants only than with *all* coding variants, suggesting that R/LF variants are more likely to be causal and may more often point directly to relevant genes, whereas common coding variants may more often be proxies for common noncoding variants that affect nearby genes.

In addition, our gene set enrichment analyses now provide supporting evidence for a role of non-neuronal mechanisms as well. Specifically, *CIDEA* and *ADH1B* are both strongly predicted to be

members of enriched gene sets related to insulin action and adipocyte biology, and both are markers that distinguish brown from white fat depots in mice[25] and humans[26]. CIDEA is predominantly expressed in adipose tissue and known as a key regulator of energy metabolism[25]. *Cidea*-deficient mice are resistance to diet-induced obesity with increased lipolysis and mitochondrial uncoupling[25]. The connection of *ADH1B* to obesity is less clear, but the gene is highly expressed in human adipocytes, has been implicated by gene expression analyses in obesity and insulin resistance, and functions early in a potentially relevant metabolic pathway (retinoid biosynthesis)[25,26,74,75]. Similar pathways were implicated by recent work dissecting the signal near *FTO*[13]. However, because SNV-association signals at *ADH1B* and *CIDEA* did not reache array-wide significance, additional genetic analysis of their role in obesity would be warranted.

In summary, we performed association analyses between R/LF variants and BMI in >700,000 individuals, and identified 14 variants in 13 genes, in 5 known and 8 novel genes. While each variant contributes little to BMI variation in the general population, they may have substantial impact on body weight at an individual level. Furthermore, prior literature for these genes and unbiased gene set enrichment analysis indicate a strong role for neuronal biology and also provide new support for a causal role of aspects of adipocyte biology. The identified genes provide potential targets that may lead to new and more precise approaches for the treatment of obesity, which has seen minimal innovation in the past 30 years[1].

URLs

CDC 2000 Growth Charts: http://www.cdc.gov/growthcharts/cdc_charts.htm

CHOP cohort: http://www.metabolic-programming.org/obesity/

EC-DEPICT code: https://github.com/RebeccaFine/obesity-ec-depict

ENSEMBL: www.ensembl.org

EasyQC: www.genepi-regensburg.de/easyqc

EasyStrata: www.genepi-regensburg.de/easystrata

ExAC: http://exac.broadinstitute.org/

GCTA: http://cnsgenomics.com/software/gcta/

GTEx: http://www.gtexportal.org/home/

GTOOL: http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html

Impute2: https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

INTERVAL Study: http://www.intervalstudy.org.uk/

PLINK v1.90: https://www.cog-genomics.org/plink2

QCTOOL: http://www.well.ox.ac.uk/~gav/qctool/#overview

RAREMETALWORKER: http://genome.sph.umich.edu/wiki/RAREMETALWORKER

RareMETALS: http://genome.sph.umich.edu/wiki/RareMETALS

RVTEST: https://github.com/zhanxw/rvtests

Shapeit2: https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html

UKHLS: https://www.understandingsociety.ac.uk/

UK10K Obesity Sample Sets - SCOOP: http://www.uk10k.org/studies/obesity.html

1000 Genomes Phase 1: http://www.1000genomes.org/category/phase-1/

ACKNOWLEDGMENTS

Alex Reiners was supported by R01DK089256. Alex Hewitt is supported by an NHMRC Practitioner Fellowship (APP1103329). Alisa Manning received funding from NIH/NIDDK K01 DK107836. Andrew Hattersley is a Wellcome Trust Senior Investigator (WT098395), and a NIH Research Senior Investigator. Andrew Morris is a Wellcome Trust Senior Fellow in Basic Biomedical Science (WT098017). Andrew Wood is supported by the European Research Council (SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC). Anne Jackson is supported by the American Heart Association (13POST16500011) and NIH (R01DK089256, R01DK101855, K99HL130580). Bratati Kahali and Elizabeth Speliotes were supported by the Doris Duke Medical Foundation, NIH (R01DK106621), the University of Michigan Internal Medicine Department, Division of Gastroenterology, the University of Michigan Biological Sciences Scholars Program and The Central Society for Clinical Research. Cristen Willer is supported by NIH (HL094535, HL109946). Daniel Liu is supported by R01HG008983 and R21DA040177. Daniel Witte is supported by the Danish Diabetes Academy, which is funded by the Novo Nordisk Foundation. Veiko Salomaa has been supported by the Finnish Foundation for Cardiovascular Research. Folkert Asselbergs is supported by a Dekker scholarship-Junior Staff Member 2014T001 Netherlands Heart Foundation and UCL Hospitals NIHR Biomedical Research Centre. Fotios Drenos is supported by the UK MRC (MC_UU_12013/1-9). Gabriela Partida received scholarship support from the University of Queensland and QIMR Berghofer. Guillaume Lettre is funded by the Montreal Heart Institute Foundation and the Canada Research Chair program. Hanieh Yaghootkar and Tim Frayling are supported by the European Research Council (323195, SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC). Iris Heid is supported by BMBF (01ER1206) and BMBF (01ER1507m), NIH and Max Planck Society. Jeff Haessler was supported by NHLBI R21HL121422. Joel Hirschhorn is supported by NIH R01DK075787. Kari North was supported by NIH (R01DK089256, R01HD057194, U01HG007416, R01DK101855), and AHA (13GRNT16490017). Manuel Rivas is supported by Nuffield Department of Clinical Medicine Award, Clarendon Scholarship. Mark McCarthy is a Wellcome Trust Senior Investigator (WT098381), and a NIH Research Senior Investigator. Mengmeng Du is supported by the NCI (R25CA94880, P30CA008748).

Pal Njolstad is supported by the European Research Council (AdG, 293574), Research Council of Norway, University of Bergen, KG Jebsen Foundation, Helse Vest, Norwegian Diabetes Association. Patrick Ellinor is supported by the NIH (1R01HL092577, R01HL128914, K24HL105780), an Established Investigator Award from the American Heart Association (13EIA14220013) and by the Foundation Leducq (14CVD01). Paul Auer was supported by NHLBI R21HL121422 and R01DK089256. Paul Huang is support by NIH (NS33335, HL57818). Rebecca Fine is supported by NIH (T32GM096911). Ruth loos is supported by the NIH (R01DK110113, U01HG007417, R01DK101855, R01DK107786). Steven Lubitz is supported by NIH (K23HL114724) and a Doris Duke Charitable Foundation Clinical Scientist Development Award. Timothy Spector holds an ERC Advanced Principal Investigator award. Trevor Mori is supported by an Australian National Health and Medical Research Fellowship (APP1042255). Tune Pers received a Lundbeck Foundation and Benzon Foundation support. Valerie Turcot is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research (CIHR). Zoltan Kutalik is supported by the Leenaards Foundation, Swiss National Science Foundation (31003A-143914) and SystemsX.ch (51RTP0_151019). Part of this work was conducted using the UK Biobank resource (Project Numbers 1251, 9072). A full list of acknowledgments appears in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Writing Group (wrote and edited manuscript)

A.E.J., A.E.L., C.M.L, C.S., G.L., H.M.H, J.N.H., K.E.N, K.L.Y., M.F.F., M. Graff, P.D., R.J.F.L., T.M.F., V.T., Y. Lu

Data preparation group (program development and quality control of data from contributing cohorts for meta-analyses)

A.E.J., A.E.L., C.S., D.J.L., E. Marouli, H.M.H., I.B.B., K.L.Y., K. Stirrups, K.S.L., M.A.R., M.C.M.G., M.C.Y.N., M. Graff, N.G.D.M., P. Mudgal, R.J.F.L., S. Feng, S.M.W., S.S., S.V., T.A., T.E., T. Karaderi, T.W.W., V.T., X.Z., Y. Lu

BMI meta-analyses (*discovery and follow-up, single-variant and gene-based*)

A.E.J., C.S., C.T.L., D.J.L., H.M.H., I.B.B., J.N.H., K.L.Y., R.J.F.L., T.M.F., V.T., Y. Lu

Childhood data (analyses and interpretation)

A.E.H., G.M., H.H., I. Barroso, I.S.F., J.P.B., S.F.A.G., V.M.

Pleiotropy working group

A.M., C.J.W., C.M.L., D.J.L., E. Marouli, F.D., G.A., G.M., G.M.P., H. Kitajima, H.M.H., J.C.F., J.P.C., J.R.B.P., J.W., K.S.R., M. Boehnke, M.I.M., P.B.M., P.D., R.J.F.L., S. Kathiresan, S.M.W., S.W., T.F.V., X.S.

Phenome-wide association studies

A.G., A.M., J.C.D., L. Bastarache, M.I.M., T.L.E.

Gene-set enrichment analyses

D.L., J.N.H., R.S.F., S.B., T.H.P., Z.K.

Monogenic and syndromic gene enrichment analyses

A.K.M., H.M.H.

Fly Obesity Screen

A. Lempradl, J.A. Pospisilik

Overseeing of contributing studies

A. Linneberg, A. Peters, A. Tönjes, A.C.H., A.D.M., A.G.U., A.H.F., A.I.d.H., A.J.L., A.M.D., A.P.M., A.P.R., A.S.B., A.T. Hattersley, A.W.H., B.B., B.G.N., C.A.B., C.C., C.E.P., C.H., C.J.W., C.L., C.M.L., C.N.A.P., D.F.E., D.F.R., D.I.C., D.M.W., D.O.M.K., D.R.N., D.R.W., D.S., D.W.B., E.B., E.B.L., E.D.A., E.F., E.I., E.K.S., E.P.B., E.Z., F. Karpe, F. Kee, F. Renström, F.W.A., G. Dedoussis, G. Tromp, G.B., G.B.J., G.K.H., G.L., G.P., G.P.J., G.W.M., H. Kuivaniemi, H.B., H.D.W., H.H., H.J.G., H.M.d.R., H.R.W., I. Barroso, I. Brandslund, I.B.B., I.F., I.J.D., I.M.H., I.R., I.S.F., J. Kaprio, J.C.C., J.C.T., J.D., J.D.R., J.F., J.G.W., J.I.R., J.M.M.H., J.M.S., J.R.O., J.S.K., J.T., K. Stefansson, K. Strauch, K.B.K., K.E.N., K.K., K.K.A., K.L.M., K.M.H., K.N., K.R.O., K.T.Z., L.E.W., L.L., L.W., M. Blüher, M. Kähönen, M. Kumari, M.A.I., M.A.R., M.B.S., M.C.H.d.G, M.d.H., M.E.J., M.F., M.H.B., M.I.M., M.L.O., M.M., M.P., M.P.D., M.S., M.U., M.V., M.W., N.D.P., N.J.S., N.J.W., N.S., N.v.L., O. Pedersen, O. Polasek, O.T.R., P. Kovacs, P.A., P.A.P., P.B.M., P.D., P.E., P.G., P.G.L., P.J.S., P.L.A., P.L.H., P.L.P., P.M.R., P.R.N., P.T.C., P.W.F., R.A.O., R.A.S., R.C., R.E.S., R.J.F.L., R.V., S. Fauser, S. Kathiresan, S.E.M., S.F.A.G., S.H. Scholte, S.J., S.L.R.K., S.M., S.P., Svati H. Shah, T.A.M., T.B.H., T.D.S., T.E., T.H., T.J., T.L., T.L.E., T.M.F., U.T., V. Gudnason, V. Salomaa, V.V., W.H.H.S., X.G., X.L., Y.Liu

Genotyping of contributing studies

A. Loukola, A. Tybjaerg-Hansen, A.B., A.D.E., A.G.U., A.I.d.H., A.J.L., A.L.M., A.M., A.M.D., A.P.G., A.P.M., A.P.R., A.R.H., A.S.B., A.V., A.W.H., A.Y.C., B.G.N., C.A.B., C.E.P., C.H., C.J.P., C.K., C.L., C.M., C.M.L., C.M.v.D., C.N.A.P., C.S., D.F.R., D.I.C., D.J.C., D.J.R., D.M.W., D.R.N., E.D.A., E.E., E.I., E.K., E.M.L., E.P.B., E.W.D., F. Karpe, F. Rivadeneira, F.S.C., G. Davies, G. Tromp, G.P.J., G.W.M, H. Kuivaniemi, H.H., H.L.G., H. Li, H.V., I.G., J. Kuusisto, J.A. Perry, J.B.J., J.C.C., J.D., J.D.F., J.G.D., J.I.R., J.M.M.H., J.S., J.S.K., J.T., K. Strauch, K.B.K., K.D.T., K.E.S., K.M., L. Milani, L. Southam, L.A. Lange, L.A.K., L.M.Y.A., L.P.L., M. Benn, M. Boehnke, M. Gorski, M. Kähönen, M. Kumari, M.B.S., M.C.H.d.G., M.C.M.G., M.F., M.H.B., M.I.M., M.L., M.L.B., M.L.G., M.L.O., M.M.N., M.P.D., M.P.S.L., N.D.P., N.G., N.J.S., N.J.W., N.v.L., O.H., P.B.M., P.G.L., P.I.W.d.B., P.T.C., P.W.F., R.A.O., R.A.S., R.E.S., R.F.S., R.J.F.L., R.L.G., R.R., R.Y., S. Kanoni, S. Kathiresan, S.C., S.F.A.G., S.F.N., S.H.V.L.B., S.L.R.K., S.S., S.W.v.d.L., T.A.L., T.B.H., T.E., T.H., T.L., U.V., V.V., W. Zhao, W.O., X.L., Y. Lu, Y.D.I.C., Y.H., Y. Liu, Y.W.

Phenotyping of contributing studies

A. Pattie, A. Peters, A.R.H., A. Robino, A.S.B., A.T. Hattersley, A. Tönjes, A. Tybjaerg-Hansen, A.U.J., A.V., A.W.H., B.B., B.G.N., B.H.T., B.K., C.A.B., C.E.L., C.E.P., C.H., C.J.P., C.K., C.M., C.M.L., C.N.A.P., C.S., D.E., D.F.R., D.I.C., D.J.R., D.R.N., D.R.V.E., D.R.W., E.C., E.D.A., E.E., E.F., E.I., E.P.B., E.R.B.P., E.T., E.W.D., F. Karpe, F. Kee, F. Renström, F. Rivadeneira, F.W., G.B.J., G.L., G.P.J., G. Tromp, G.W.M., H.B., H.D.W., H.H., H.L.G., H.Li, I.J.D., I.R., J.C., J.C.C., J.D., J.D.F., J.F., J.H.J., J. Kaprio, J. Lindström, J.M.M.H., J.M.S., J.P.B., J.S., J.S.K., K.B.K., K.E.N, K.H.L., K.K., K.K.A., K.M.H., K.N., K.R.O., K.S.L., K.S.S., K.T.Z., L.A.K., L.E.B., L.E.W., L.L., L.M.Y.A., L. Southam, L. Sun, L.W., M.A., M.A.I., M. Blüher, M. Brumat, M.C.H.d.G., M.C.M.G., M.F.F., M.I.M., M.J.C., M. Kähönen, M. Karaleftheri, M. Kumari, M.L.B., M.M., M.M.N., M.N., M.R., M.S., N.D.P., N.G., N.J.S., N.J.W., N.L., N.N., N.R.L., O.H., O.H.F., O.L.H., O. Polasek, O.R., O.T.R., P.A., P.G.L., P. Komulainen,

P. Kovacs, P.L.P., P.M.R., P. Mitchell, P.R.K., P.R.N., P.T.C., P.T.E., R.d.M., R.E.S., R.F.S., R.M.C., R.R., R.S.K., R.V., R.Y., S.A.L., S.E.M., S.F.A.G., S. Fauser, S.H. Scholte, S.H. Shah, S.H.V., S.L.R.K., S.M., S.S., S.T., T.A.L., T.A.M., T.B.H., T.D.S., T.E.G., T.J., T.J.P., T.L., T.L.E., T.N.P., V. Giedraitis, V. Salomaa, V.T., W.H.H.S., W.O., X.L., X.S., Y. Liu, Y. Lu

Data analysis of contributing studies

A.E.H., A.E.J., A.E.L., A.G., A.J.C., A.J.S., A. Lophatananon, A.M., A.P.M., A.P.P., A. Pirie, A.R.W., A. Rasheed, A. Robino, A.S.B., A. Teumer, A.V.S., A.Y.C., B.K., C.A.B., C.A.W., C.H., C.P.N., C.S., C.T.H., C.T.L., D.F.R., D.I.C., D.J.T., D.M.W., D.S.A., D.V., E.B.W., E.E., F. Rivadeneira, G.C.P., G. Davies, G.L., G.M., G. Thorleifsson, G. Tromp, G.V., H. Li, H. Lin, H.M.S., H.P., H.R.W., H.T., H.Y., I.F., I.G., J.A. Perry, J.B.J., J.C.C., J.C.G., J.E.H., J.G., J.G.D., J.H.Z., J. Haessler, J. Hernesniemi, J.I.R., J. Kuusisto, J. Li, J. Luan, J.M.M.H., J.P.B., J.P.T., J.R.O., J.S.K., J.v.S., J.W.J., J.Y., K.B.K., K.E.N., K.E.S., K.E.T., K.L.Y., K.M., K.S.L., L.-A. Lin, L.A. Lange, L. Broer, L.F.B., L.H., L.M.O.L., L.M.Y.A., L. Moilanen, L.P.L., L. Southam, M.A.N., M.C., M.C.H.d.G., M.C.M.G., M.C.Y.N., M.D., M.F., M.F.F., M. Gorski, M. Graff, M. Kumari, M.L., M.P.S.L., M.R., M.U., M.V., N.D.P., N.G., N.G.D.M., N.J.S., N.L.H.C., N.R.R., N.v.L., O.H., P.L.A., P. Mudgal, P.S., P.Y., R.A.S., R.C., R.L.G., R.U., R.Y., S.A.L., S.E.A., S.G., S.J., S.M., S.M.W., S.P., S.S., S.V., S.W.v.d.L., T.E., T. Karaderi, T. Korhonen, T.L.E., T.M.F., T.S.C., T.V.V., T.W.W., V.M., V. Steinthorsdottir, V.T., V.V., W.G., W. Zhang, W. Zhou, W. Zhao, X.G., X.L., X.S., Y.H., Y.J., Y. Lu, Y.S., Y. Wu

REFERENCES

1. Bray, G.A. & Ryan, D.H. Update on obesity pharmacotherapy. *Ann N Y Acad Sci* **1311**, 1-13 (2014).
2. Bray, G.A., Fruhbeck, G., Ryan, D.H. & Wilding, J.P. Management of obesity. *Lancet* **387**, 1947-56 (2016).
3. Monda, K.L. *et al.* A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat Genet* **45**, 690-6 (2013).
4. Wen, W. *et al.* Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum Mol Genet* **23**, 5492-504 (2014).
5. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
6. Winkler, T.W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* **11**, e1005378 (2015).
7. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* **49**, 1458-1467 (2017).
8. van der Klaauw, A.A. & Farooqi, I.S. The hunger genes: pathways to obesity. *Cell* **161**, 119-32 (2015).
9. Edwards, S.L., Beesley, J., French, J.D. & Dunning, A.M. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet* **93**, 779-97 (2013).
10. Stratigopoulos, G. *et al.* Hypomorphism of Fto and Rpgrip1l causes obesity in mice. *J Clin Invest* **126**, 1897-910 (2016).
11. Stratigopoulos, G., LeDuc, C.A., Cremona, M.L., Chung, W.K. & Leibel, R.L. Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated and retinitis pigmentosa GTPase regulator-interacting protein-1-like (RPGRIP1L) genes and coordinates leptin receptor signaling. *J Biol Chem* **286**, 2155-70 (2011).
12. Stratigopoulos, G. *et al.* Hypomorphism for RPGRIP1L, a ciliary gene vicinal to the FTO locus, causes increased adiposity in mice. *Cell Metab* **19**, 767-79 (2014).
13. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* **373**, 895-907 (2015).
14. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371-5 (2014).
15. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* (2017).
16. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-7 (2016).
17. Sina, M. *et al.* Phenotypes in three pedigrees with autosomal dominant obesity caused by haploinsufficiency mutations in the melanocortin-4 receptor gene. *Am J Hum Genet* **65**, 1501-7 (1999).
18. Pearce, L.R. *et al.* KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* **155**, 765-77 (2013).
19. Hinney, A. *et al.* Several mutations in the melanocortin-4 receptor gene including a nonsense and a frameshift mutation associated with dominantly inherited obesity in humans. *J Clin Endocrinol Metab* **84**, 1483-6 (1999).
20. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-90 (2014).
21. van den Berg, L. *et al.* Melanocortin-4 receptor gene mutations in a Dutch cohort of obese children. *Obesity (Silver Spring)* **19**, 604-11 (2011).
22. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-1161 (2016).
23. Lunetta, K.L. *et al.* Rare coding variants and X-linked loci associated with age at menarche. *Nat Commun* **6**, 7756 (2015).
24. GIANT. Include reference to GIANT WHR paper.

25. Zhou, Z. *et al.* Cidea-deficient mice have lean phenotype and are resistant to obesity. *Nat Genet* **35**, 49-56 (2003).
26. Tews, D. *et al.* Comparative gene array analysis of progenitor cells from human paired deep neck and subcutaneous adipose tissue. *Mol Cell Endocrinol* **395**, 41-50 (2014).
27. Stutzmann, F. *et al.* Non-synonymous polymorphisms in melanocortin-4 receptor protect against obesity: the two facets of a Janus obesity gene. *Hum Mol Genet* **16**, 1837-44 (2007).
28. Lin, H.Q., Wang, Y., Chan, K.L., Ip, T.M. & Wan, C.C. Differential regulation of lipid metabolism genes in the brain of acetylcholinesterase knockout mice. *J Mol Neurosci* **53**, 397-408 (2014).
29. Vignaud, A. *et al.* Genetic ablation of acetylcholinesterase alters muscle function in mice. *Chem Biol Interact* **175**, 129-30 (2008).
30. Ji, Z., Mei, F.C. & Cheng, X. Epac, not PKA catalytic subunit, is required for 3T3-L1 preadipocyte differentiation. *Front Biosci (Elite Ed)* **2**, 392-8 (2010).
31. Yan, J. *et al.* Enhanced leptin sensitivity, reduced adiposity, and improved glucose homeostasis in mice lacking exchange protein directly activated by cyclic AMP isoform 1. *Mol Cell Biol* **33**, 918-26 (2013).
32. Almahariq, M., Mei, F.C. & Cheng, X. Cyclic AMP sensor EPAC proteins and energy homeostasis. *Trends Endocrinol Metab* **25**, 60-71 (2014).
33. Kai, A.K. *et al.* Exchange protein activated by cAMP 1 (Epac1)-deficient mice develop beta-cell dysfunction and metabolic syndrome. *FASEB J* **27**, 4122-35 (2013).
34. Hardie, D.G., Ross, F.A. & Hawley, S.A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat Rev Mol Cell Biol* **13**, 251-62 (2012).
35. Hardie, D.G. & Ashford, M.L. AMPK: regulating energy balance at the cellular and whole body levels. *Physiology (Bethesda)* **29**, 99-107 (2014).
36. Lopez, M., Nogueiras, R., Tena-Sempere, M. & Dieguez, C. Hypothalamic AMPK: a canonical regulator of whole-body energy balance. *Nat Rev Endocrinol* **12**, 421-32 (2016).
37. Minokoshi, Y. *et al.* AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428**, 569-74 (2004).
38. Viollet, B. *et al.* The AMP-activated protein kinase alpha2 catalytic subunit controls whole-body insulin sensitivity. *J Clin Invest* **111**, 91-8 (2003).
39. Xue, B. *et al.* Neuronal protein tyrosine phosphatase 1B deficiency results in inhibition of hypothalamic AMPK and isoform-specific activation of AMPK in peripheral tissues. *Mol Cell Biol* **29**, 4563-73 (2009).
40. Warren, H.R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* (2017).
41. Chami, N. *et al.* Exome Genotyping Identifies Pleiotropic Variants Associated with Red Blood Cell Traits. *Am J Hum Genet* **99**, 8-21 (2016).
42. Li, M. *et al.* SOS2 and ACP1 Loci Identified through Large-Scale Exome Chip Analysis Regulate Kidney Development and Function. *J Am Soc Nephrol* (2016).
43. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
44. Schwartz, M.W., Woods, S.C., Porte, D., Jr., Seeley, R.J. & Baskin, D.G. Central nervous system control of food intake. *Nature* **404** 661-671 (2000).
45. Garfield, A.S. *et al.* A neural basis for melanocortin-4 receptor-regulated appetite. *Nat Neurosci* **18**, 863-71 (2015).
46. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88** 131-141 (1997).
47. Fan, W., Boston, B.A., Kesterson, R.A., Hruby, V.j. & Cone, R.D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385** 165-168 (1997).
48. Yeo, G.S.H. *et al.* A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genetics* **20** 111-112 (1998).

49. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nature Genetics* **20** 113-114 (1998).
50. Farooqi, I.S. *et al.* Clinical Spectrum of Obesity and Mutations in the Melanocortin 4 Receptor Gene. *The New England Journal of Medicine* **348**, 1085-1095 (2003).
51. Lubrano-Berthelier, C. *et al.* Melanocortin 4 receptor mutations in a large cohort of severely obese adults: prevalence, functional classification, genotype-phenotype relationship, and lack of association with binge eating. *J Clin Endocrinol Metab* **91**, 1811-8 (2006).
52. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
53. Hinney, A. *et al.* Melanocortin-4 receptor gene: case-control study and transmission disequilibrium test confirm that functionally relevant mutations are compatible with a major gene effect for extreme obesity. *J Clin Endocrinol Metab* **88**, 4258-67 (2003).
54. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* **42**, 142-8 (2010).
55. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-948 (2010).
56. Miyawaki, K. *et al.* Inhibition of gastric inhibitory polypeptide signaling prevents obesity. *Nat Med* **8**, 738-742 (2002).
57. Hansotia, T. *et al.* Extrapancreatic incretin receptors modulate glucose homeostasis, body weight, and energy expenditure. *J Clin Invest* **117**, 143-52 (2007).
58. Fulurija, A. *et al.* Vaccination against GIP for the treatment of obesity. *PLoS One* **3**, e3163 (2008).
59. Finan, B. *et al.* Reappraisal of GIP Pharmacology for Metabolic Diseases. *Trends Mol Med* **22**, 359-76 (2016).
60. Irwin, N. & Flatt, P.R. Evidence for beneficial effects of compromised gastric inhibitory polypeptide action in obesity-related diabetes and possible therapeutic implications. *Diabetologia* **52**, 1724-31 (2009).
61. Revelli, J.P. *et al.* Profound obesity secondary to hyperphagia in mice lacking kinase suppressor of ras 2. *Obesity (Silver Spring)* **19**, 1010-8 (2011).
62. Costanzo-Garvey, D.L. *et al.* KSR2 is an essential regulator of AMP kinase, energy expenditure, and insulin sensitivity. *Cell Metab* **10**, 366-78 (2009).
63. Brommage, R. *et al.* High-throughput screening of mouse knockout lines identifies true lean and obese phenotypes. *Obesity (Silver Spring)* **16**, 2362-7 (2008).
64. Liu, L. *et al.* Proteomic characterization of the dynamic KSR-2 interactome, a signaling scaffold complex in MAPK pathway. *Biochim Biophys Acta* **1794**, 1485-95 (2009).
65. Kuhnen, P. *et al.* Proopiomelanocortin Deficiency Treated with a Melanocortin-4 Receptor Agonist. *N Engl J Med* **375**, 240-6 (2016).
66. Beiroa, D. *et al.* GLP-1 agonism stimulates brown adipose tissue thermogenesis and browning through hypothalamic AMPK. *Diabetes* **63**, 3346-58 (2014).
67. Xiang, Y.Y., Dong, H., Yang, B.B., Macdonald, J.F. & Lu, W.Y. Interaction of acetylcholinesterase with neurexin-1beta regulates glutamatergic synaptic stability in hippocampal neurons. *Mol Brain* **7**, 15 (2014).
68. Bartels, C.F., Zelinski, T. & Lockridge, O. Mutation at codon 322 in the human acetylcholinesterase (ACHE) gene accounts for YT blood group polymorphism. *Am J Hum Genet* **52**, 928-36 (1993).
69. Farlow, M.R. *et al.* Effectiveness and tolerability of high-dose (23 mg/d) versus standard-dose (10 mg/d) donepezil in moderate to severe Alzheimer's disease: A 24-week, randomized, double-blind study. *Clin Ther* **32**, 1234-51 (2010).
70. Farlow, M. *et al.* Safety and tolerability of donepezil 23 mg in moderate to severe Alzheimer's disease. *BMC Neurol* **11**, 57 (2011).
71. Tariot, P., Salloway, S., Yardley, J., Mackell, J. & Moline, M. Long-term safety and tolerability of donepezil 23 mg in patients with moderate to severe Alzheimer's disease. *BMC Res Notes* **5**, 283 (2012).

72. Hu, Y. *et al.* Role of Exchange Protein Directly Activated by Cyclic AMP Isoform 1 in Energy Homeostasis: Regulation of Leptin Expression and Secretion in White Adipose Tissue. *Mol Cell Biol* **36**, 2440-50 (2016).
73. Altarejos, J.Y. *et al.* The Creb1 coactivator Crtc1 is required for energy balance and fertility. *Nat Med* **14**, 1112-7 (2008).
74. Winnier, D.A. *et al.* Transcriptomic identification of ADH1B as a novel candidate gene for obesity and insulin resistance in human adipose tissue in Mexican Americans from the Veterans Administration Genetic Epidemiology Study (VAGES). *PLoS One* **10**, e0119941 (2015).
75. Molotkov, A., Deltour, L., Foglio, M.H., Cuenca, A.E. & Duester, G. Distinct retinoid metabolic functions for alcohol dehydrogenase genes Adh1 and Adh4 in protection against vitamin A toxicity or deficiency revealed in double null mutant mice. *J Biol Chem* **277**, 13804-11 (2002).
76. G. TEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
77. Volpicelli-Daley, L.A., Duysen, E.G., Lockridge, O. & Levey, A.I. Altered hippocampal muscarinic receptors in acetylcholinesterase-deficient mice. *Ann Neurol* **53**, 788-96 (2003).
78. Ivanenkov, V.V., Murphy-Piedmonte, D.M. & Kirley, T.L. Bacterial expression, characterization, and disulfide bond determination of soluble human NTPDase6 (CD39L2) nucleotidase: implications for structure and function. *Biochemistry* **42**, 11726-35 (2003).
79. Jain, R.N. *et al.* Hip1r is expressed in gastric parietal cells and is required for tubulovesicle formation and cell survival in mice. *J Clin Invest* **118**, 2459-70 (2008).
80. Engqvist-Goldstein, A.E. *et al.* RNAi-mediated Hip1R silencing results in stable association between the endocytic machinery and the actin assembly machinery. *Mol Biol Cell* **15**, 1666-79 (2004).
81. Tao, Y.X. The melanocortin-4 receptor: physiology, pharmacology, and pathophysiology. *Endocr Rev* **31**, 506-43 (2010).
82. Farooqi, I.S. *et al.* Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**, 1085-95 (2003).
83. Stutzmann, F. *et al.* Prevalence of melanocortin-4 receptor deficiency in Europeans and their age-dependent penetrance in multigenerational pedigrees. *Diabetes* **57**, 2511-8 (2008).
84. Vaisse, C. *et al.* Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity. *J Clin Invest* **106**, 253-62 (2000).
85. Schonke, M., Myers, M.G., Jr., Zierath, J.R. & Bjornholm, M. Skeletal muscle AMP-activated protein kinase gamma1(H151R) overexpression enhances whole body energy homeostasis and insulin sensitivity. *Am J Physiol Endocrinol Metab* **309**, E679-90 (2015).
86. Pellinen, T. *et al.* Small GTPase Rab21 regulates cell adhesion and controls endosomal traffic of beta1-integrins. *J Cell Biol* **173**, 767-80 (2006).
87. Banerjee, U. & Cheng, X. Exchange protein directly activated by cAMP encoded by the mammalian rapgef3 gene: Structure, function and therapeutics. *Gene* **570**, 157-67 (2015).
88. Rippey, C. *et al.* Formation of chimeric genes by copy-number variation as a mutational mechanism in schizophrenia. *Am J Hum Genet* **93**, 697-710 (2013).
89. Schmitz, C., Kinge, P. & Hutter, H. Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive Caenorhabditis elegans strain nre-1(hd20) lin-15b(hd126). *Proc Natl Acad Sci U S A* **104**, 834-9 (2007).
90. Setoguchi, R. *et al.* Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. *Science* **319**, 822-5 (2008).
91. Widom, R.L., Culic, I., Lee, J.Y. & Korn, J.H. Cloning and characterization of hcKrox, a transcriptional regulator of extracellular matrix gene expression. *Gene* **198**, 407-20 (1997).
92. Sun, X. *et al.* Deletion of atbf1/zfhx3 in mouse prostate causes neoplastic lesions, likely by attenuation of membrane and secretory proteins and multiple signaling pathways. *Neoplasia* **16**, 377-89 (2014).

93. Parsons, M.J. *et al.* The Regulatory Factor ZFHX3 Modifies Circadian Function in SCN via an AT Motif-Driven Axis. *Cell* **162**, 607-21 (2015).
94. Balzani, E. *et al.* The Zfhx3-Mediated Axis Regulates Sleep and Interval Timing in Mice. *Cell Rep* **16**, 615-21 (2016).
95. Kao, Y.H. *et al.* ZFHX3 knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes. *Int J Cardiol* **210**, 85-92 (2016).

FIGURE LEGENDS

Figure 1. Effect sizes (y-axis) of the 14 BMI-associated R/LF coding variants by their minor allele frequency. Effect sizes are expressed in body weight (kg) per allele, assuming a SD of 4.5 kg and an average-sized person of 1.7m tall. Solid markers indicate that the minor allele is associated with higher BMI, and clear markers indicate that the minor allele is associated with lower BMI. Variants were identified in all-ancestry analyses (light blue diamonds), the European ancestry analyses (dark blue square) and women-only analyses (pink diamond). Effect sizes for previously identified GWAS loci are shown in navy blue diamonds. The dotted line represents 80% power, assuming $\alpha = 2x10^{-7}$ and N= 525,000 (discovery sample size).

Figure 2. Heatmap showing DEPICT gene set enrichment results for suggestive and significant rare and low-frequency coding SNVs. For any given square, the color indicates how strongly the corresponding gene (x-axis) is predicted to belong to the reconstituted gene set (y-axis), based on the gene's Z-score for gene set inclusion in DEPICT's reconstituted gene sets (red indicates a higher, blue a lower Z-score). To visually reduce redundancy and 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**). Heatmap intensity and DEPICT *P*-values (**Supplementary Table 17**) correspond to the most significantly enriched gene set within the meta-gene set. Annotations for genes indicate (1) whether it has an OMIM annotation as underlying a monogenic obesity disorder (black/grey), (2) the MAF of the significant ExomeChip (EC) variant (blue), (3) whether the variant's *P*-value reached array-wide significance ($<2x10^{-7}$) or suggestive significance ($<5x10^{-4}$) (purple), (4) whether the variant was novel, overlapping "relaxed" GWAS signals from Locke *et al.*[5] (GWAS $P<5x10^{-4}$), or overlapping "stringent" GWAS hits (GWAS $P<5x10^{-8}$) (pink), and (5) whether the gene was included in the gene set enrichment analysis or excluded by filters (orange/brown) (**Online Methods, Supplementary Note**). Annotations for gene sets indicate if the meta-gene set was significant (green, FDR <0.01, <0.05, or not significant) in the DEPICT analysis of GWAS results[5]. Here, two regions of particularly strong gene set

membership are shown (see full heat map in **Supplementary Figure 10a)**.

TABLES

Table 1 Rare and low-frequency coding variants significantly associated with BMI

Chr:position	Variant	Coding locus	Allele		Amino acid change	EAF (%)	β (SD/allele)	SE	*P*-value	*N*	Explained variance (%)
			Effect	Other							
All-ancestries additive											
1:154987704	rs141845046	*ZBTB7B**	T	C	p.Pro190Ser	2.44%	0.048	0.006	7.73E-18	718,628	0.011%
7:100490797	rs1799805	*ACHE**	T	G	p.His353Asn	3.90%	0.029	0.005	2.82E-10	707,448	0.006%
12:48143315	rs145878042	*RAPGEF3**	G	A	p.Leu300Pro	1.10%	0.066	0.008	1.56E-15	700,852	0.010%
12:49399132	rs1126930	*PRKAG1*	C	G	p.Thr38Ser	3.22%	0.034	0.005	3.98E-12	712,354	0.007%
12:72179446	rs61754230	*RAB21**	T	C	p.Ser224Phe	1.74%	0.040	0.007	1.33E-09	693,373	0.005%
12:117977550	rs56214831	*KSR2*	T	C	p.Arg525Gln	0.82%	0.057	0.010	1.08E-08	655,049	0.005%
12:123345509	rs34149579	*HIP1R*	T	G	p.Cys938Phe	4.54%	-0.032	0.004	2.00E-14	716,253	0.009%
16:72830539	rs62051555	*ZFHX3**	G	C	p.Gln1100His	4.34%	-0.024	0.004	4.01E-08	690,637	0.005%
18:58039478	rs13447324	*MC4R*	T	G	p.Tyr35Ter	0.01%	0.542	0.086	2.26E-10	631,683	0.006%
19:46178020	rs139215588	*GIPR*	A	G	p.Arg190Gln	0.11%	-0.148	0.028	1.25E-07	695,800	0.005%
19:46180976	rs143430880	*GIPR*	G	A	p.Glu288Gly	0.13%	-0.153	0.028	2.96E-08	599,574	0.006%
20:25195509	rs6050446	*ENTPD6**	A	G	p.Lys185Glu	2.71%	-0.034	0.005	2.40E-10	717,084	0.006%
All-ancestries sex-specific additive (women only)											
19:3813906	rs45465594	*ZFR2**	C	A	p.Ile718Met	2.55%	-0.040	0.008	1.94E-07	373,848	0.008%
European Ancestry additive											
9:97062981	rs12236219	*ZNF169**	T	C	p.Arg381Cys	4.23%	-0.029	0.005	8.78E-10	612,396	0.007%

Array-wide significant is defined as $P < 2x10^{-7}$.

Variant positions are reported according to Build 37 and their alleles are coded based on the positive strand.

Alleles (effect/other), effect allele frequency (EAF), beta (b), standard error (SE) and P values are based on the meta-analysis of Discovery Stage (GIANT) and Validations stage (deCODE, UKBiobank) studies. Effect allele is always the minor allele. Effects (b) are expressed in SD, assuming mean=0 and SD=1.

The amino acid change from the most abundant coding transcript is shown in this table (see Supplementary Table 25 for more details on protein annotation based on VEP tool and transcript abundance from GTEx database).

* Novel gene, i.e. not previously implicated in human obesity

Table 2. Genes significantly associated with BMI in a gene-based meta-analyses, aggregating R/LF coding SNVs

Gene	Location longest coding transcript	Test[d]	*N* variants	*P*-value	*Conditioned P*-value[a]	Single variant	
						Top variant	*P*-value
All-ancestries sex-combined							
SLC6A17	chr1:110693132-110744823	SKAT	13	2.73E-07	0.13	rs41313405	4.45E-07
RAPGEF3	chr12:48128453-48152889	SKAT	19	8.91E-15	0.20	rs145878042	5.16E-14
PRKAG1	chr12:49396055-49412629	SKAT	4	2.75E-12	0.53	rs1126930	2.63E-12
RAB21	chr12:72148643-72187256	SKAT	5	4.81E-08	0.27	rs61754230	4.96E-08
KSR2	chr12:117890817-118406028	SKAT	7	7.15E-09	0.19	rs56214831	4.59E-08
MAP1A	chr15:43809806-43823818	SKAT	25	9.42E-07	0.16	rs55707100	1.01E-06
MC4R	chr18:58038564-58040001	VT	4	3.72E-09	0.01	rs13447325	2.97E-11
GIPR	chr19:46171502-46186982	VT	10	8.24E-09	1.12E-04	rs143430880	5.76E-06
All-ancestries sex-specific							
ALDH3A1 (men only)	chr17:19641298-19651746	SKAT	15	3.24E-07	0.003	rs142078447	8.62E-06
ZFR2 (women only)	chr19:3804022-3869027	SKAT	19	1.81E-07	0.82	rs45465594	3.64E-07
European sex-combined							
ACHE	chr7:100487615-100493592	SKAT	6	3.30E-10	0.12	rs386545548	7.22E-10
European sex-specific							
ANGPTL7 (men only)	chr1:11249346-11256038	VT	3	2.50E-06	0.008	rs202182115	2.56E-05
ZNF169 (women only)	chr9:97021548-97064111	SKAT	9	1.89E-07	0.24	rs12236219	1.06E-06

Array-wide significant gene-based association is defined as $P<2.5x10^{-6}$. P-values are based on the meta-analysis of Discovery Stage studies.

Gene-based analyses were performed with SKAT and VT, results shown are from the test (SKAT or VT) for which the significance exceeded $P<2.5x10^{-6}$. Only results using the "broad" SNV inclusion criteria reached array-wide significance.

Transcript positions are reported according to Build 37 for the longest coding transcript supported by RefSeq (as displayed in USCS Genome Browser).

[a]P-value after conditioning on the most significant (top) single variant aggregated in the gene-based test.

BOX 1 – Brief description of the 13 genes (alphabetical) identified

ACHE (acetylcholinesterase). *ACHE* is mainly expressed in brain and muscle[76]. Its encoded protein hydrolyzes acetylcholine (Ach) at brain cholinergic synapses and neuromuscular junctions, and thus terminates signal transmission[67]. Knockout mice showed a reduction in expression of muscarinic Ach receptors in brain regions associated with learning and memory and showed lower ability to initiate the signaling cascade[77]. This gene has fewer missense variants than expected and is highly intolerant to loss of function (LoF) mutations[52].

ENTPD6 (ectonucleoside triphosphate diphosphohydrolase 6). Previously known as *Interleukin 6 Signal Transducer-2*, this gene is similar to E-type nucleotidases that participate in purine and pyrimidine metabolism, calcium ion binding, hydrolase activity, magnesium ion binding and nucleoside-diphosphatase activity[78]. It is widely expressed in many different tissues, in particular in the brain[76].

GIPR (gastric inhibitory polypeptide receptor). *GIPR* encodes a G-protein coupled receptor for gastric inhibitory polypeptide that is secreted by intestinal K-cells after food ingestion[59]. GIPR activation stimulates insulin secretion from pancreatic β-cells and mediates fat deposition by increasing lipoprotein lipase activity, lipogenesis, fatty acid and glucose uptake in adipocytes. GIPR is mostly expressed in EBV-transformed lymphocytes, stomach and visceral adipose tissue[76].

HIP1R (huntingtin interacting protein 1 related). HIP1R is a multi-domain protein that promotes actin binding and cell survival and interacts with CLTB and HIP1 (GeneCards). HIP1 and HIP1R appear to play central roles in clathrin-coated vesicle formation and intracellular membrane trafficking by promoting transient interaction between actin filaments and the endocytic machinery[79,80]. HIP1R is most expressed in the stomach tissue, brain (substantia nigra, spinal cord, hippocampus), and sun-exposed skin [76].

KSR2 (kinase suppressor of ras 2). KSR2 is an intracellular protein that functions as a molecular scaffold to regulate MAP kinases ERK1/2 and determine cell fates. KSR2 also regulates AMPK activity

controlling cellular thermogenesis, fat oxidation, and glucose metabolism[18,61,62]. Knockout mouse models and human mutations have been linked to obesity risk[62]. KSR2 is almost exclusively expressed in the brain. It has fewer missense variants than expected and is highly intolerant to LoF mutations[52].

MC4R (melanocortin 4 receptor). MC4R is a seven-transmembrane G-protein coupled receptor, predominantly expressed in the brain[76]. MC4R has been known to play a key role in body weight regulation for more than 20 years. Activation of MC4R by α-MSH, a POMC-derived peptide, suppresses food intake, MC4R antagonists increase food intake and *MC4R* deficiency in human and rodent models results in hyperphagia and severe and early-onset obesity[81]. More than 150 *MC4R* mutations have been identified in individuals with severe, early-onset obesity[81], many of which lead to a complete or partial loss of function[82,83]. Up to 6% of individuals with severe, early-onset obesity carry pathogenic mutations in *MC4R*, making *MC4R* deficiency the most common form of monogenic obesity[82,84].

PRKAG1 (protein kinase AMP-activated non-catalytic subunit gamma 1). The protein encoded by *PRKAG1* is one of the gamma regulatory subunits of the AMP-activated protein kinase (AMPK), which is an important energy-sensing enzyme that monitors cellular energy status[34]. AMPK and PRKAG1 are ubiquitously expressed[76]. In the hypothalamus, AMPK influences food intake, energy expenditure and glucose homeostasis[36]. Muscle-specific overexpression of AMPK γ1 subunit in mice results in increased food intake, but does not affect body weight, presumably through a compensatory increased energy expenditure[85].

RAB21 (member RAS oncogene family). RAB21 belongs to the Rab family of monomeric GTPases involved in the control of cellular membrane traffic. The encoded protein is widely expressed[76] and plays a role in the targeted trafficking of integrins, and is involved in the regulation of cell adhesion and migration[86]. *RAB21* is thought to be intolerant to LoF mutations[52].

RAPGEF3 (rap guanine nucleotide exchange factor 3, also EPAC1). *RAPGEF3* encodes the exchange protein directly activated by cAMP isoform 1 (EPAC1), one of two cAMP sensors that are involved in numerous intracellular cAMP-mediated functions[87]. EPAC1 is ubiquitously expressed[76], and insights from

mouse knockout models suggest a role in energy homeostasis and the development of obesity and diabetes through the regulation of leptin and insulin signaling[31,87].

ZFR2 (zinc finger RNA binding protein 2). The biological function of the gene product is as yet undetermined. GO annotations related to this gene include *nucleic acid binding*. It may have a role in dendritic branching and axon guidance[88,89]. ZFR2 is predominantly expressed in the brain[76].

ZBTB7B (zinc finger and BTB domain containing 7B, also ThPOK). ZBTB7B is a transcription factor regulating T-cell fate in the thymus, particularly as the master regulator of $CD4^+$ lineage commitment[90]. It is a repressor of type 1 collagen gene expression[91]. This gene is mainly expressed in T-cell lineages, skin and gastrointestinal tissues. *ZBTB7B* is thought to be intolerant to LoF mutations[52].

ZFHX3 (zinc finger homeobox 3). *ZFHX3* encodes a transcription factor with multiple homeodomains and zinc finger motifs and plays a role in cell-cycle, myogenic and neuronal differentiation. This gene is a tumor suppressor[92] that influences circadian rhythms[93,94] and sleep[94]. It may also contribute to the genesis of atrial fibrillation[95]. *ZFHX3* is highly expressed in arterial tissue and also other tissues[76]. The *ZFHX3* gene is highly intolerant to LoF mutations[52].

ZNF169 (zinc finger protein 169). The biological function of the gene product is as yet unclear. GO annotations suggest that *ZNF169* is involved in *nucleic acid binding and transcriptional regulation*. This gene is ubiquitously expressed[76].

More details and references in Supplementary Table 24.

ONLINE METHODS

Study design & participants

The discovery cohort consisted of 123 studies (163 datasets) comprising 526,508 adult (≥18yrs) individuals of the following ancestries (**Supplementary Figure 1**): 1) European (N = 449,889), 2) South Asian (N = 29,398), 3) African (N = 27,610), 4) East Asian (N = 8,839), and 5) Hispanic (N = 10,772). All participating institutions and coordinating centers approved this project and informed consent was obtained from all study participants. Discovery meta-analyses were carried out in each ancestry separately and in the All-ancestries combined group, for both sex-specific and sex-combined analyses. SNVs for which associations reach suggestive significance ($P<2.0x10^{-6}$) in the discovery analyses, were taken forward for follow-up in 192,226 individuals of European ancestry from the UK BioBank and deCODE. Conditional analyses were conducted in the All-ancestries and European descent groups. Study-specific design, sample quality control and descriptive statistics are provided in **Supplementary Tables 1-3**.

Phenotype

Body mass index (BMI: weight [in kilograms] / height [in meters] ^{2}) was corrected for age, age^2 and genomic principal components (PC, derived from GWAS data, the variants with MAF > 1% on ExomeChip, or ancestry informative markers available on the ExomeChip), as well as any additional study-specific covariates (e.g. recruiting center), in a linear regression model. For studies with non-related individuals, residuals were calculated separately by sex, whereas for family-based studies sex was included as a covariate in the model. Additionally, residuals for case/control studies were calculated separately. Finally, residuals were subject to inverse normal transformation[96].

Genotype calling

The majority of studies followed a standardized protocol and performed genotype calling using the designated manufacturer software, which was then followed by zCall[97]. For 10 studies, participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the raw intensity data for the samples from seven genotyping centers were assembled into a single project for joint

calling[98]. Study-specific quality control (QC) measures of the genotyped variants were implemented before association analysis (**Supplementary Table 2**).

Statistical analyses

Study-level association analyses. Individual cohorts were analyzed separately for each ancestry, in sex-combined and sex-specific groups, with either RAREMETALWORKER (see URL links at the end of the **Online Methods**) or RVTEST[99] (**Supplementary Table 2**), to associate inverse normal transformed BMI with genotype accounting for potential cryptic relatedness (kinship matrix) in a linear mixed model. These software tools are designed to perform score-statistics based rare-variant association analyses, can accommodate both unrelated and related individuals, and provide single-variant results and variance-covariance matrices. The covariance matrix captures linkage disequilibrium (LD) relationships between markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses[100]. Single-variant analyses were performed for both additive and recessive models.

Centralized quality-control. A centralized quality-control procedure, implemented in EasyQC[101], was applied to individual cohort association summary statistics to identify cohort-specific problems: (1) assessment of possible problems in BMI transformation, (2) comparison of allele frequency alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues, and (3) examination of quantile-quantile (QQ) plots per study to identify any problems arising from population stratification, cryptic relatedness and genotype biases.

Meta-analyses. Meta-analyses were carried out by two different analysts at two sites in parallel. We excluded variants with a call rate < 95%, Hardy-Weinberg equilibrium P-value < 1×10^{-7}, or large allele frequency deviations from reference populations (> 0.6 for all-ancestry analyses and > 0.3 for ancestry-specific population analyses). Significance for single-variant analyses was defined at the array-wide level (a Bonferroni-corrected threshold of $P < 2\times10^{-7}$ for ~250,000 SNVs). To test for sex-differences of the significant variants ($P < 2\times10^{-7}$), we calculated the P-diff for each SNP, which tests for differences between women-specific and men-specific beta estimates using EasyStrata[102]. For gene-based analyses,

we applied the sequence kernel association test (SKAT)[103] and the Variable Threshold (VT)[104] gene-based methods using two different sets of criteria (broad and strict) to select predicted damaging R/LF variants with MAF < 5%, based on coding variant annotation from five prediction algorithms (PolyPhen2 HumDiv and HumVar, LRT, MutationTaster and SIFT)[20]. Our *broad* gene-based tests included nonsense, stop-loss, splice site, and missense variants that are annotated as damaging by at least one algorithm mentioned above. Our *strict* gene-based tests included only nonsense, stop-loss, splice site, and missense variants annotated as damaging by all five algorithms. Statistical significance for gene-based tests was set at a Bonferroni-corrected threshold of $P < 2.5\times10^{-6}$ for about 20,000 genes[16,105]. Singe-variant and gene-based meta-analyses were both performed using RareMETALS R-package[106]. As our secondary analyses are nested and/or highly correlated with our primary analysis, we chose the same, already stringent, Bonferroni-corrected significance threshold for both analyses.

Genomic inflation. Although the overall λ_{GC} value is in the normal range for all coding variants (λ_{GC} = 1.1, **Supplementary Table 23**), we observed a marked genomic inflation of the test statistics even after adequate control for population stratification (linear mixed model) arising from common markers (λ_{GC} = 1.99, **Supplementary Figure 2a** and **Supplementary Table 23**). Such inflation is expected for a highly polygenic trait like BMI, as was previously confirmed for height[15], and is consistent with our very large sample size[5,107]. Furthermore, some of the inflation may be due to the design of the ExomeChip, which besides R/LF coding SNVs also contains (common and non-coding) SNVs that include previously identified GWAS loci for all traits, including for BMI and BMI-related traits, reported in the GWAS catalogue at the time of its design.

After removing established loci (+/- 1Mb), the excess of significant associations is markedly reduced and inflation reduced (**Supplementary Figures 2c and 2d**).

Furthermore, to exclude the possibility that some of the observed associations between BMI and R/LF SNVs could be due to allele calling problems in the smaller studies, we performed a sensitivity meta-analysis with primarily European ancestry studies totaling >5,000 participants. We found very concordant

effect sizes, suggesting that smaller studies do not bias our results (**Supplementary Figure 12**).

Follow-up Analysis. We sought additional evidence for association of the top signals ($P<2.0x10^{-6}$) identified in the discovery meta-analysis using two independent studies from the UK (UK Biobank, interim release, N = 119,613) and Iceland (deCODE, N = 72,613), respectively (**Supplementary Tables 1-3)**. We used the same QC and analytical methodology as described above. We used the inverse-variance weighted fixed effects meta-analysis in METAL[108], to combine the discovery and follow-up association results. Significant associations were defined at $P < 2\times10^{-7}$ in the combined meta-analysis of discovery, UK Biobank and deCODE results.

***Effect of study design*.** To investigate the potential effect of study design of the participating studies, we tested for heterogeneity between population-based, all case-control studies, T2D case-control studies (**Supplementary Table 26**). None of these comparisons showed significant evidence of heterogeneity ($P<7.4x10^{-5}$, correcting for multiple testing).

Conditional analyses. The RareMETALS R-package[106] was used to identify independent BMI associated signals across the all-ancestry meta-analysis results in the discovery phase. RareMETALS performs conditional analyses by using covariance matrices from each individual cohort to distinguish true signals from the shadows of adjacent significant variants in LD. The conditional associations of all the variants within 1Mb of each R/LF coding variant were analyzed to identify [1] nearby secondary signals and [2] to determine independence from nearby non-coding variants or previously identified GWAS loci (previously defined as a window of 1Mb surrounding the lead SNP). Gene-based conditional analyses were also performed in RareMETALS.

Due to the selective coverage of variants on the ExomeChip, we also conducted the respective conditional analyses in the UK Biobank dataset that included 847,441 genome-wide genotyped markers, and 72,355,667 variants imputed against UK10k haplotype reference panel, merged with the 1000 Genomes Phase 3 reference panel. Where available, directly genotyped variants where used for conditional analyses. Otherwise, imputed variants with good imputation quality (IMPUTE2 info score >

0.6) were used. We used QCTOOL to extract variants of interest from the original imputed data set. Subsequently, GTOOL was used to convert to PLINK format (genotype calling threshold 0.99) and merged with the directly genotyped variants for conditional analyses in PLINK v1.90b3.35 64-bit (25 Mar 2016).

Conversions of effect size and explained variants. We assumed that 1 SD = 4.5 kg/m^2 BMI-units, based on population based data, and 1.7m as the average height of a person to convert effects sizes in SD-units into body weight. The variance explained by each variant was calculated using the effect allele frequency (f) and beta (β) from the meta analyses using the formula[109] of explained variance = $2f(1-f)\beta^2$.

Penetrance analysis. We examined the penetrance for the four rare SNVs, p.Arg525Gln (rs56214831) in *KSR2*, p.Tyr35Ter (rs13447324) in *MC4R*, and p.Arg190Gln (rs139215588) and p.Glu288Gly (rs143430880) in *GIPR* in European ancestry data from the UKBiobank (N up to 120,000). For each variant, we compared the prevalence of underweight (BMI < 18.5 kg/m^2), normal weight (18.5 kg/m^2 ≤ BMI < 25 kg/m^2), overweight (25 kg/m^2 ≤ BMI < 30 kg/m^2) and obesity (BMI ≥ 30 kg/m^2) of non-carriers with non-carriers. We used a Pearson χ^2 test to test for difference between distributions, and a χ^2 for linear trend to test whether distributions of carriers were shifted compared to non-carriers. For p.Arg525Gln in *KSR2* and p.Tyr35Ter in *MC4R*, we hypothesized that obesity prevalence was higher in carriers than in non-carriers, whereas for the two GIPR variants, we hypothesized that the prevalence of normal weight was higher in carriers than non-carriers.

Associations with obesity for the coding rare and low-frequency loci in children. For each of the 14 R/LF SNVs, we tested for association with childhood obesity in the CHOP cohort (Childhood Obesity: Early Programming by Infant Nutrition), the Severe Childhood Onset Obesity Project (SCOOP), the UK Household Longitudinal Study (UKHLS) and INTERVAL Study (INTERVAL). Summary statistics across the studies were combined using a fixed effects inverse-variance meta-analysis with METAL[108].

In the CHOP study, cases (1,358 boys, 1,060 girls) were defined as having a BMI > 95th percentile at any point in their childhood. Controls (1,412 boys, 1,143 girls) were defined as having < 50th percentile consistently through throughout childhood. The BMI percentiles are based on the CDC 2000 Growth Charts. All children were classified based on their BMI measurements between the ages of 2 and 18. All individuals are of European ancestry and were collected at the Children's Hospital of Philadelphia. Informed consent was obtained from all study participants and study protocols were approved by the local ethics committees. Genotypes were obtained using the HumanHap550v1, HumanHap550v3, and Human610-Quad high-density SNP arrays from Illumina. The intersection of all SNPs on the arrays was used in all subsequent pre-imputation analyses. Before imputation, we excluded SNPs with a Hardy-Weinberg equilibrium *P*-value < 1.0×10^{-6}, call rate of < 95% or MAF of < 1%. The genotypes were then pre-phased using Shapeit2 and imputed using the 1000 Genomes Phase 1 integrated variant set with Impute2. After imputation, SNPs were excluded if the INFO score was < 0.4. Boys and girls were analyzed separately using a logistic regression of case and control status, adjusting for three eigenvectors, and summary statistics were combined using a fixed effects inverse-variance meta-analysis with METAL[108].

SCOOP is a sub-cohort of the Genetics Of Obesity Study (GOOS) cohort. It includes >1,500 UK European ancestry individuals with severe, early onset obesity (BMI Standard Deviation Score > 3 and obesity onset before the age of 10 years), in whom known monogenic causes of obesity have been excluded (cases with *MC4R* mutations were excluded). Two case-control analyses with SCOOP cases were performed: 1) SCOOP vs. UKHLS for which array (Illumina HumanCoreExome) data was available, and 2) SCOOP vs. INTERVAL, for whom whole-exome sequencing data was available.

For the array based analyses, UKHLS controls were genotyped on the Illumina HumanCoreExome-12v1-0 Beadchip. SCOOP cases and 48 UKHLS controls were genotyped on the Illumina HumanCoreExome-12v1-1 Beadchip. The 48 overlapping UKHLS samples were used for quality control to ensure there were no systematic differences and bias between the two versions of the chip. SCOOP and UKHLS samples were phased with SHAPEITv2, and imputed with IMPUTE2 using

the combined UK10K-1000G Phase III reference panel. For the WES analyses, SCOOP vs. INTERVAL controls were WES within the UK10K-EXOME project (Agilent v3) and the INTERVAL project (Agilent v5) respectively and were then jointly called and QC-ed on the union of the sequencing baits. Individuals overlapping or related between the array based and WES studies were removed.

After QC, 1,456 SCOOP and 6,460 UKHLS (BMI range 19-30), and 521 SCOOP and 4,057 INTERVAL individuals were available for the two analyses, all were unrelated, of high quality, and of European ancestry. For both analyses (i.e. SCOOP vs. UKHLS and SCOOP vs. INTERVAL), a maximum likelihood frequentist association test with the additive genetic model was implemented in SNPTEST v2.5. In the SCOOP vs. UKHLS analysis, sex and the first six PCs were included as covariates and variants with a SNPTEST INFO score <0.4 and HWE $p<10^{-6}$ were removed. For the SCOOP vs INTERVAL analysis, we performed an unadjusted analysis (adjustment for PCs did not change sufficiently the results) and variants were limited to those covered at ≥7x in at least 80% of each sequencing cohort, meeting the VQSR threshold of –2.52, missingness <80%, HWE *P*-value$<10^{-8}$, and GQ ≥30.

Cross-trait analyses. We evaluated each of the 14 R/LF SNVs for their association with other relevant obesity-related traits and conditions. We performed lookups in ExomeChip meta-analysis results from other consortia, including, our own GIANT consortium (height[15], WHR adjusted for BMI[24]), MAGIC (HbA1c, Fasting Insulin, Fasting Glucose, 2-hour glucose), GLGC (HDL-cholesterol (HDL-C), LDL-cholesterol (LDL-C), triglycerides and total cholesterol)), IBPC[40] (systolic and diastolic blood pressure), REPROGEN[23] (age at menarche and menopause) and GoT2D/T2D-GENES[16] (type 2 Diabetes). Associations were considered significant at $P < 2.0\times10^{-5}$, accounting for multiple testing.

Phenome-wide association analysis (PheWAS). To evaluate the potential for pleiotropic effects for SNPs discovered from primary analyses, we performed phenome-wide association studies (PheWASs) using genotype and phenotype data from two independent sources of electronic health records (EHR): Vanderbilt University Medical Center Biorepository (BioVU) and the United Kingdom BioBank

(UKBB). Phenotype selection and analysis strategy were synchronized across sites. A total of 1502 hierarchical phenotype codes from EHRs were curated by grouping International Classification of Disease, Ninth Revision (ICD-9) clinical/billing codes as previously described[110]. Phenotype codes with 20 or more cases and with minor allele count of 5 or greater in cases and controls were eligible for analysis. Series of logistic regression analyses were then performed in individuals of European ancestry for each eligible phenotype-genotype combination while adjusting for 5 genetic ancestry PCs. Odds ratios from genotype-phenotype combinations present in both BioVU and UKBB were then aggregated using inverse-variance weighted fixed-effects meta-analysis. Associations with p-values corresponding to false discovery rate (FDR) cut off of less than 10% were considered statistically significant.

Gene set enrichment analysis. We adapted DEPICT, a gene set enrichment analysis method for GWAS data, for use with the ExomeChip ('EC-DEPICT'). DEPICT's primary innovation is the use of "reconstituted" gene sets, where many different types of gene sets (e.g. canonical pathways, protein-protein interaction networks, and mouse phenotypes) were extended through the use of large-scale microarray data (see[111] for details). EC-DEPICT computes *P*-values based on Swedish ExomeChip data (Malmö Diet and Cancer [MDC], All New Diabetics in Scania [ANDIS], and Scania Diabetes Registry [SDR] cohorts, N=11,899) and, unlike DEPICT, takes as input only coding variants and only the genes directly containing those variants, rather than all genes within a specified amount of linkage disequilibrium (**Supplementary Note**).

Four analyses were performed for the BMI EC variants: [1] all coding variants with $P<5x10^{-4}$, [2] all coding variants with $P<5x10^{-4}$ independent of known GWAS variants[5], [3] all coding R/LF variants with $P<5x10^{-4}$, and [4] all coding R/LF variants with $P<5x10^{-4}$ independent of known GWAS variants. Affinity propagation clustering[3] was used to group highly correlated gene sets into "meta-gene sets". For each meta-gene set, the member gene set with the best *P*-value was used as representative for purposes of visualization (**Supplementary Note**). DEPICT for ExomeChip was written using the Python programming language (See URLs).

Drosophila RNAi knockdown experiments. For each of the 13 genes in which R/LF coding variants were associated with BMI, we searched for its corresponding orthologues in *Drosophila* in the ENSEMBL orthologue database. Orthologues were available for nine genes, but missing for *ZBTB7B, MC4R, GIPR,* and *ZNF169*. For each of the nine genes, we generated adipose-tissue (cg-Gal4) and neuronal (elav-Gal4) specific RNAi-knockdown crosses, leveraging upstream activation sequence (UAS)-inducible short-hairpin knockdown lines, available through the Vienna Drosophila Resource Center (VDRC). We crossed male UAS-RNAi flies and elav-GAL4 or CG-GAL4 virgin female flies. All fly experiments were carried out at 25 °C. Five-to-seven-day-old males were sorted into groups of 20, weighed and homogenated in PBS with 0,05% Tween with Lysing Matrix D in a beadshaker. The homogenate was heat-inactivated for 10 min in a thermocycler at 70 °C. 10µl of the homogenate was subsequently used in triglyceride assay (Sigma, Serum Triglyceride Determination Kit) which was carried out in duplicates according to protocol, with one alteration: the samples were cleared of residual particulate debris by centrifugation before absorbance reading. Resulting triglyceride values were normalized to fly weight and larval/population density. We used the non-parametric Kruskall-Wallis test to compare wild type with knockdown lines.

Enrichment analysis in monogenic genes of obesity. We identified 39 genes with strong evidence that disruption causes monogenic or syndromic forms of obesity (**Supplementary Table 21**). To test whether these genes are enriched for R/LF coding variant associations with BMI, we conducted simulations by matching each of the 39 genes with other genes based on gene length and number of variants tested, to create a matched set of genes. We generated 1,000 matched gene sets from our data and assessed how often the number of R/LF coding variants that exceeded given significance thresholds was greater in our monogenic/syndromic obesity gene set compared to the matched gene sets.

DATA AVAILABILITY

Summary statistics can be downloaded from http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium

REFERENCES

96. Auer, P.L., Reiner, A.P. & Leal, S.M. The effect of phenotypic outliers and non-normality on rare-variant association testing. *Eur J Hum Genet* **24**, 1188-94 (2016).
97. Goldstein, J.I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543-5 (2012).
98. Grove, M.L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8**, e68095 (2013).
99. Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-6 (2016).
100. Liu, D.J. *et al.* Meta-analysis of gene-level tests for rare variant association. *Nat Genet* **46**, 200-4 (2014).
101. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-212 (2014).
102. Winkler, T.W. *et al.* EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259-61 (2015).
103. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).
104. Price, A.L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* **86**, 832-8 (2010).
105. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nat Genet* **44**, 623-30 (2012).
106. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828-9 (2014).
107. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**, 807-12 (2011).
108. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
109. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* **41**, 18-24 (2009).
110. Denny, J.C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205-10 (2010).
111. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).