Genome-Wide Meta-Analysis of 241,258 Adults Accounting for Smoking Behavior Identifies Novel Loci for Obesity Traits

Anne E Justice*[1], Thomas W Winkler*[2], Mary F Feitosa*[3], Misa Graff*[1], Virginia A Fisher*[4], Kristin Young*[1], Llilda Barata*[3], Xuan Deng[4], Jacek Czajkowski[3], David Hadley[5, 6], Julius S Ngwa[7, 4], Tarunveer S Ahluwalia[8, 9], Audrey Y Chu[10, 11], Nancy L Heard-Costa[12, 10], Elise Lim[4], Jeremiah Perez[4], John D Eicher[13], Zoltán Kutalik[14-16], Luting Xue[4], Anubha Mahajan[17], Frida Renström[18, 19], Joseph Wu[4], Qibin Qi[20], Shafqat Ahmad[19], Tamuno Alfred[21, 22], Najaf Amin[23], Lawrence F Bielak[24], Amelie Bonnefond[25], Jennifer Bragg[26, 27], Gemma Cadby[28], Martina Chittani[29], Scott Coggeshall[30], Tanguy Corre[14-16], Nese Direk[31, 32], Joel Eriksson[33], Krista Fischer[34], Mathias Gorski[2, 35], Marie Neergaard Harder[8], Momoko Horikoshi[17, 36], Tao Huang[37, 38], Jennifer E Huffman[13, 39], Anne U Jackson[27], Johanne Marie Justesen[8], Stavroula Kanoni[40], Leena Kinnunen[41], Marcus E Kleber[42], Pirjo Komulainen[43], Meena Kumari[44, 45], Unhee Lim[46], Jian'an Luan[47], Leo-Pekka Lyytikäinen[48, 49], Massimo Mangino[50, 51], Ani Manichaikul[52], Jonathan Marten[39], Rita PS Middelberg[53], Martina Müller-Nurasyid[54-56], Pau Navarro[39], Louis Pérusse[57, 58], Natalia Pervjakova[34, 59], Cinzia Sarti[60], Albert Vernon Smith[61, 62], Jennifer A Smith[24], Alena Stančáková[63], Rona J Strawbridge[64, 65], Heather M Stringham[27], Yun Ju Sung[66], Toshiko Tanaka[67], Alexander Teumer[68], Stella Trompet[69, 70], Sander W van der Laan[71], Peter J van der Most[72], Jana V Van Vliet-Ostaptchouk[73], Sailaja L Vedantam[74, 75], Niek Verweij[76], Jacqueline M Vink[77, 78], Veronique Vitart[39], Ying Wu[79], Loic Yengo[25], Weihua Zhang[80, 81], Jing Hua Zhao[47], Martina E Zimmermann[2], Niha Zubair[82], Gonçalo R Abecasis[27], Linda S Adair[83], Saima Afaq[80, 81], Uzma Afzal[80, 81], Stephan JL Bakker[84], Traci M Bartz[30, 85], John Beilby[86-88], Richard N Bergman[89], Sven Bergmann[15, 16], Reiner Biffar[90], John Blangero[91], Eric Boerwinkle[92], Lori L Bonnycastle[93], Erwin Bottinger[21, 94], Daniele Braga[29], Brendan M Buckley[95], Steve Buyske[96, 97], Harry Campbell[98], John C Chambers[81, 80, 99], Francis S Collins[93], Joanne E Curran[91], Gert J de Borst[100], Anton JM de Craen[+70], Eco JC de Geus[77, 101], George Dedoussis[102], Graciela E Delgado[42], Hester M den Ruijter[71], Gudny Eiriksdottir[61],

Anna Eriksson[33], Tõnu Esko[34], Jessica D Faul[103], Ian Ford[104], Terrence Forrester[105], Karl Gertow[64, 65], Bruna Gigante[106], Nicola Glorioso[107], Jian Gong[82], Harald Grallert[108-110], Tanja B Grammer[42], Niels Grarup[8], Saskia Haitjema[71], Göran Hallmans[111], Anders Hamsten[64, 65], Torben Hansen[8], Tamara B Harris[112], Catharina A Hartman[113], Maija Hassinen[43], Nicholas D Hastie[39], Andrew C Heath[114], Dena Hernandez[115], Lucia Hindorff[116], Lynne J Hocking[117, 118], Mette Hollensted[8], Oddgeir L Holmen[119], Georg Homuth[120], Jouke Jan Hottenga[77], Jie Huang[121], Joseph Hung[122, 123], Nina Hutri-Kähönen[124, 125], Erik Ingelsson[126-128], Alan L James[86, 129, 122], John-Olov Jansson[130], Marjo-Riitta Jarvelin[131-134], Min A Jhun[24], Marit E Jørgensen[9], Markus Juonala[135, 136], Mika Kähönen[137, 138], Magnus Karlsson[139], Heikki A Koistinen[41, 140, 141], Ivana Kolcic[142], Genovefa Kolovou[143], Charles Kooperberg[82], Bernhard K Krämer[42], Johanna Kuusisto[144], Kirsti Kvaløy[145], Timo A Lakka[146, 43], Claudia Langenberg[47], Lenore J Launer[112], Karin Leander[106], Nanette R Lee[147, 148], Lars Lind[149], Cecilia M Lindgren[150, 17], Allan Linneberg[151-153], Stephane Lobbens[25], Marie Loh[80], Mattias Lorentzon[33], Robert Luben[154], Gitta Lubke[155], Anja Ludolph-Donislawski[54, 156], Sara Lupoli[29], Pamela AF Madden[114], Reija Männikkö[43], Pedro Marques-Vidal[157], Nicholas G Martin[53], Colin A McKenzie[105], Barbara McKnight[30, 85, 158], Dan Mellström[33], Cristina Menni[50], Grant W Montgomery[159], AW (Bill) Musk[86, 160, 161], Narisu Narisu[93], Matthias Nauck[162], Ilja M Nolte[72], Albertine J Oldehinkel[113], Matthias Olden[2], Ken K Ong[47], Sandosh Padmanabhan[163, 118], Patricia A Peyser[24], Charlotta Pisinger[164, 165], David J Porteous[166, 118], Olli T Raitakari[167, 168], Tuomo Rankinen[169], DC Rao[66, 114, 170], Laura J Rasmussen-Torvik[171], Rajesh Rawal[108, 109], Treva Rice[66, 114], Paul M Ridker[11, 172], Lynda M Rose[11], Stephanie Rosse[82], Igor Rudan[98], Serena Sanna[173], Mark A Sarzynski[169], Naveed Sattar[174], Kai Savonen[43], David Schlessinger[175], Salome Scholtens[72], Claudia Schurmann[21, 22], Robert A Scott[47], Bengt Sennblad[64, 65, 176], Marten A Siemelink[71], Günther Silbernagel[177], P Eline Slagboom[178], Harold Snieder[72], Jan A Staessen[179, 180], David J Stott[181], Morris A Swertz[182], Amy J Swift[93], Kent D Taylor[183, 184], Bamidele O Tayo[185], Barbara Thorand[109, 110], Dorothee Thuillier[25], Jaakko Tuomilehto[186-189], Andre G Uitterlinden[190, 31], Liesbeth Vandenput[33], Marie-Claude Vohl[58, 191], Henry Völzke[68], Judith M Vonk[72], Gérard Waeber[157], Melanie Waldenberger[108,

[109], RGJ Westendorp[192], Sarah Wild[98], Gonneke Willemsen[77], Bruce HR Wolffenbuttel[73], Andrew Wong[193], Alan F Wright[39], Wei Zhao[24], M Carola Zillikens[190], Damiano Baldassarre[194, 195], Beverley Balkau[196], Stefania Bandinelli[197], Carsten A Böger[35], Dorret I Boomsma[77], Claude Bouchard[169], Marcel Bruinenberg[198], Daniel I Chasman[11, 199], Yii-Der Ida Chen[200], Peter S Chines[93], Richard S Cooper[185], Francesco Cucca[173, 201], Daniele Cusi[202], Ulf de Faire[106], Luigi Ferrucci[67], Paul W Franks[19, 203, 38], Philippe Froguel[25, 204], Penny Gordon-Larsen[83, 205], Hans-Jörgen Grabe[206, 207], Vilmundur Gudnason[61, 62], Christopher A Haiman[208], Caroline Hayward[39, 118], Kristian Hveem[145], Andrew D Johnson[13], J Wouter Jukema[69, 209, 210], Sharon LR Kardia[24], Mika Kivimaki[45], Jaspal S Kooner[211, 81, 99], Diana Kuh[193], Markku Laakso[144], Terho Lehtimäki[48, 49], Loic Le Marchand[46], Winfried März[212, 213], Mark I McCarthy[36, 17, 214], Andres Metspalu[34], Andrew P Morris[215, 17], Claes Ohlsson[33], Lyle J Palmer[216], Gerard Pasterkamp[71, 217], Oluf Pedersen[8], Annette Peters[109, 110], Ulrike Peters[82], Ozren Polasek[142, 98], Bruce M Psaty[218-220], Lu Qi[38], Rainer Rauramaa[43, 221], Blair H Smith[222, 118], Thorkild IA Sørensen[8, 223, 224], Konstantin Strauch[54, 156], Henning Tiemeier[225], Elena Tremoli[194, 195], Pim van der Harst[76, 182, 226], Henrik Vestergaard[8, 9], Peter Vollenweider[157], Nicholas J Wareham[47], David R Weir[103], John B Whitfield[53], James F Wilson[227, 39], Jessica Tyrrell[228, 229], Timothy M Frayling[230], Inês Barroso[231-233], Michael Boehnke[27], Panagiotis Deloukas[40, 231, 234], Caroline S Fox[10], Joel N Hirschhorn[74, 235, 75], David J Hunter[236, 38, 237, 75], Tim D Spector[50], David P Strachan [5, 238], Cornelia M van Duijn[23, 239, 240], Iris M Heid[2, 241], Karen L Mohlke[79], Jonathan Marchini[242], Ruth JF Loos[†21, 22, 47, 243, 244], Tuomas O Kilpeläinen[†8, 47, 245], Ching-Ti Liu[†4], Ingrid B Borecki[†3], Kari E North[†1], L Adrienne Cupples[†4, 10]

*These authors contributed equally to this work.

[†]These authors jointly supervised this work.

[+] Anton JM de Craen recently passed away while this work was in process.

73

74

AFFILIATIONS

1. Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599.
2. Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, D-93053 Regensburg, Germany.
3. Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine; St. Louis, MO, 63108 USA.
4. Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118.
5. Population Health Research Institute, St. George's, University of London, London, SW17 0RE, UK.
6. TransMed Systems, Inc., Cupertino, CA 95014.
7. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD.
8. The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
9. Steno Diabetes Center, Gentofte, Denmark.
10. NHLBI Framingham Heart Study, Framingham, MA, 01702 USA.
11. Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA USA.
12. Department of Neurology, Boston University School of Medicine, Boston, MA, 02118, USA.
13. Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, The Framingham Heart Study, Framingham, MA, USA.
14. Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland.
15. Department of Computational Biology, University of Lausanne, Lausanne, Switzerland.

16. Swiss instititute of Bioinformatics.
17. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.
18. Department of Biobank Research, Umeå University, Umeå, Sweden.
19. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Skåne University Hospital Malmö, SE-205 02, Malmö, Sweden.
20. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA;.
21. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
22. The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
23. Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, 3015GE, The Netherlands.
24. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA.
25. University of Lille, CNRS, Institut Pasteur of Lille, UMR 8199 - EGID, Lille, France.
26. Internal Medicine - Nephrology, University of Michigan, Ann Arbor, Michigan, USA.
27. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA.
28. Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley, Australia.
29. Dept. Health Sciences, University of Milan, Via A. Di Rudiní, 8 20142, Milano, Italy.
30. Department of Biostatistics, University of Washington, Seattle, WA 98195.
31. Department of Epidemiology, Erasmus Medical Center, Rotterdam, Netherlands.

32. Department of Psychiatry, Dokuz Eylul University, Izmir, Turkey.
33. Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden.
34. Estonian Genome Center, University of Tartu, Tartu 51010, Estonia.
35. Department of Nephrology, University Hospital Regensburg, Regensburg, Germany.
36. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, OX3 7LJ, UK.
37. Epidemiology Domain, Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549.
38. Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA.
39. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland.
40. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
41. Department of Health, National Institute for Health and Welfare, Helsinki, FI-00271 Finland.
42. Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany.
43. Kuopio Research Institute of Exercise Medicine, Kuopio, Finland.
44. ISER, University of Essex, Colchester, Essex, UK CO43SQ.
45. Department of Epidemiology and Public Health, UCL, London, UK. WC1E 6BT.
46. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA.
47. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, CB2 0QQ, UK.

48. Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland.
49. Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere 33014, Finland.
50. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.
51. NIHR Biomedical Research Centre at Guy's and St. Thomas' Foundation Trust, London, UK.
52. Center for Public Health Genomics and Biostatistics Section, Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia 22903.
53. Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
54. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany.
55. Department of Medicine I, Ludwig-Maximilians-Universität, D-81377 Munich, Germany.
56. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany.
57. Department of Kinesiology, Faculty of Medicine, Université Laval, Québec, Canada.
58. Institute of Nutrition and Functional Foods, Université Laval, Québec, Canada.
59. Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia.
60. Department of Social and Health Care, City of Helsinki, Helsinki, Finland.
61. Icelandic Heart Association, Kopavogur, Iceland.
62. Faculty of Medicine, University of Iceland, Reykjavik, Iceland.
63. Department of Medicine, Institute of Clinical Medicine, University of Eastern Finland, 70210 Kuopio, Finland.
64. Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.

65. Center for Molecular Medicine, Karolinska University Hospital Solna, Stockholm, Sweden.
66. Division of Biostatistics, Washington University School of Medicine, St Louis, MO.
67. Translational Gerontology Branch, National Institute on Aging, Baltimore MD, USA.
68. Institute for Community Medicine, University Medicine Greifswald, Germany.
69. Department of Cardiology, Leiden University Medical Center, The Netherlands.
70. Department of Gerontology and Geriatrics, Leiden University Medical Center, The Netherlands.
71. Laboratory of Experimental Cardiology, Department of Cardiology, Division Heart & Lungs, UMC Utrecht, the Netherlands.
72. Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands.
73. Department of Endocrinology, University of Groningen, University Medical Center Groningen, the Netherlands.
74. Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston MA 02115 USA.
75. Broad Institute of Harvard and MIT, Cambridge, MA 02142 USA.
76. Department of Cardiology, University Medical Center Groningen, University of Groningen, the Netherlands.
77. Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands.
78. NCA, Vrije Universiteit & Vrije Universiteit Medical Center, Amsterdam, the Netherlands.
79. Department of Genetics, University of North Carolina, Chapel Hill, NC 27599 USA.
80. Dept Epidemiology and Biostatistics, School of Public Health, Imperical College London, UK.
81. Cardiology, Ealing Hospital NHS Trust, Middlesex, UK.
82. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA USA.

83. Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill NC 27599.
84. Department of Medicine, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.
85. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101.
86. Busselton Population Medical Research Institute, Nedlands, WA 6009, Australia.
87. PathWest Laboratory Medicine of WA, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
88. School of Pathology and Laboraty Medicine, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia.
89. Diabetes and Obesity Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA.
90. Clinic for Prosthetic Dentistry, Gerostomatology and Material Science, University Medicine Greifswald, Germany.
91. South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX.
92. Human Genetics Center, The University of Texas Health Science Center, PO Box 20186, Houston, TC 77225.
93. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA.
94. Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
95. Department of Pharmacology and Therapeutics, University College Cork, Ireland.

96. Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA.
97. Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854, USA;.
98. Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Scotland, UK
99. Imperial College Healthcare NHS Trust, London, UK.
100. Department of Vascular Surgery, Division of Surgical Specialties, UMC Utrecht, the Netherlands.
101. EMGO+ Institute Vrije Universiteit & Vrije Universiteit Medical Center.
102. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece.
103. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA.
104. Robertson Center for Biostatistics, University of Glasgow, United Kingdom.
105. Tropical Metabolism Research Unit, Tropical Medicine Research Institute, University of the West Indies, Mona, JMAAW15 Jamaica.
106. Unit of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
107. Hypertension and Related Disease Centre, AOU-University of Sassari.
108. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764 Neuherberg, Germany.
109. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, D-85764 Neuherberg, Germany.
110. German Center for Diabetes Research, D-85764 Neuherberg, Germany.
111. Department of Public Health and Clinical Medicine, Section for Nutritional Research, Umeå University, Umeå, Sweden.

112. Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, National Institutes of Health, Bethesda, MD.
113. Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands.
114. Department of Psychiatry, Washington University School of Medicine, St. Louis, MO.
115. Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA.
116. Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892.
117. Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK, AB25 2ZD.
118. Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, Scotland.
119. St. Olav Hospital, Trondheim University Hospital, Trondheim, Norway.
120. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Germany.
121. Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.
122. School of Medicine and Pharmacology, The University of Western Australia, 25 Stirling Hwy, Crawley, WA 6009, Australia.
123. Department of Cardiovascular Medicine, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
124. Department of Pediatrics, Tampere University Hospital, Tampere 33521, Finland.
125. Department of Pediatrics, University of Tampere School of Medicine, Tampere 33014, Finland.
126. Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala, 751 85, Sweden.

127. Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
128. Science for Life Laboratory, Uppsala University, Uppsala, 750 85, Sweden.
129. Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
130. Department of Physiology, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden.
131. Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, UK
132. Center for Life Course Epidemiology, Faculty of Medicine, P.O.Box 5000, FI-90014 University of Oulu, Finland
133. Biocenter Oulu, University of Oulu, Finland.
134. Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O.Box 20, FI-90220 Oulu, 90029 OYS, Finland.
135. Department of Medicine, University of Turku, Turku 20520 Finland.
136. Division of Medicine, Turku University Hospital, Turku 20521, Finland.
137. Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland.
138. Department of Clinical Physiology, University of Tampere School of Medicine, Tampere 33014, Finland.
139. Clinical and Molecular Osteoporosis Research Unit, Department of Orthopedics and Clinical Sciences, Skåne University Hospital, Lund University, Malmö, Sweden.
140. Department of Medicine and Abdominal Center: Endocrinology, University of Helsinki and Helsinki University Central Hospital, Helsinki, FI-00029 Finland.
141. Minerva Foundation Institute for Medical Research, Biomedicum 2U, Helsinki, FI-00290 Finland.

142. Department of Public Health, Faculty of Medicine, University of Split, Croatia.
143. Department of Cardiology, Onassis Cardiac Surgery Center, Athens, Greece.
144. Department of Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland.
145. HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway.
146. Institute of Biomedicine/Physiology, University of Eastern Finland, Kuopio Campus, Finland.
147. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City 6000, Philippines.
148. Department of Anthropology, Sociology and History, University of San Carlos, Cebu City 6000, Philippines.
149. Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala 751 85, Sweden.
150. Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford OX3 7BN, UK.
151. Research Centre for Prevention and Health, the Capital Region of Denmark, Copenhagen, Denmark.
152. Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark.
153. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
154. Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
155. Department of Psychology, University of Notre Dame, Notre Dame, USA.
156. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, D-81377 Munich, Germany.

157. Department of Medicine, Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland.
158. Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, WA 98109.
159. Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia.
160. School of Population Health, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia.
161. Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia.
162. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Germany.
163. Institute of Cardiovascular and Medical Sciences, BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Scotland.
164. Research Center for Prevention and Health, Glostrup Hospital, Glostrup Denmark.
165. Department of Public Health, Faculty of Health Sciences, University of Copenhagen, Denmark.
166. Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh.
167. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20521, Finland.
168. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland.
169. Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA.
170. Department of Genetics, Washington University School of Medicine, St. Louis, MO.

171. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL USA.
172. Division of Cardiology, Brigham and Women's Hospital, Boston MA USA.
173. Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale Delle Ricerche (CNR), Cittadella Universitaria di Monserrato, 09042, Monserrato, Italy.
174. BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, United Kingdom.
175. Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA.
176. Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden.
177. Division of Angiology, Department of Internal Medicine, Medical University of Graz, Austria.
178. Department of Molecular Epidemiology, Leiden University Medical Center, Leiden , The Netherlands.
179. Research Unit Hypertension and Cardiovascular Epidemiology, Department of Cardiovascular Science , University of Leuven, Campus Sint Rafael, Kapucijnenvoer 35, Leuven; Belgium.
180. R&D VitaK Group, Maastricht University, Brains Unlimited Building, Oxfordlaan 55, Maastricht, The Netherlands.
181. Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow, United Kingdom.
182. Department of Genetics, University of Groningen, University Medical Center Groningen, the Netherlands.
183. Center for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor/UCLA Medical Center, Torrance, CA, USA.
184. Department of Pediatrics, University of California Los Angeles, Los Angeles, CA

185. Department of Public Health Sciences, Stritch School of Medicine, Loyola University of Chicago, Maywood, IL 61053 USA.
186. Research Division, Dasman Diabetes Institute, Dasman, Kuwait
187. Department of Neurosciences and Preventive Medicine, Danube-University Krems, 3500 Krems, Austria.
188. Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland.
189. Saudi Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia.
190. Department of Internal Medicine, Erasmus Medical Center, Rotterdam, Netherlands.
191. School of Nutrition, Université Laval, Québec, Canada.
192. Department of Public Health, and Center for Healthy Ageing, University of Copenhagen, Denmark.
193. MRC Unit for Lifelong Health and Ageing at UCL, 33 Bedford Place, London, WC1B 5JU, UK.
194. Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano, Milan , Italy.
195. Centro Cardiologico Monzino, IRCCS, Milan, Italy.
196. Inserm U-1018, CESP, 94807 Villejuif cedex France.
197. Geriatric Unit, Azienda USL Toscana centro, Florence Italy.
198. Lifelines Cohort Study, PO Box 30001, 9700 RB Groningen, the Netherlands.
199. Division of Genetics, Brigham and Women's Hospital, Boston MA USA.
200. Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute and Department of Pediatrics, Harbor-UCLA, Torrance, CA 90502, USA.
201. Dipartimento di Scienze Biomediche, Universita' degli Studi di Sassari, Sassari, Italy.
202. Sanipedia srl, Bresso (Milano), Italy and Institute of Biomedical Technologies National Centre of Research Segrate (Milano), Italy.
203. Department of Public Health & Clinical Medicine, Umeå University, Umeå, Sweden.

204. Department of Genomics of Common Disease, Imperial College London, London, UK.
205. Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill NC 27516.
206. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Germany.
207. German Center for Neurodegenerative Diseases (DZNE), Site Rostock/ Greifswald, Germany.
208. Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089.
209. Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands.
210. Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands.
211. Faculty of Med, National Heart & Lung Institute, Cardiovascular Science, Hammersmith Campus, Hammersmith Hospital, Hammersmith Campus, Imperial College London, UK.
212. Synlab Academy, Synlab Services GmbH, Mannheim, Germany.
213. Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria.
214. Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Oxford, UK.
215. Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK.
216. School of Public Health, University of Adelaide, Adelaide, Australia.
217. Laboratory of Clinical Chemistry and Hematology, Division Laboratories & Pharmacy, UMC Utrecht, the Netherlands.
218. Department of Medicine, University of Washington, Seattle, WA 98195.
219. Department of Epidemiology, University of Washington, Seattle, WA 98101.
220. Group Health Research Institute, Group Health Cooperative, Seattle, WA 98101.
221. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland.

222. Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee, DD2 4RB.
223. Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospital (2000 Frederiksberg), The Capital Region, Copenhagen, Denmark.
224. MRC Integrative Epidemiology Unit, Bristol University, Bristol, UK.
225. Department of Psychiatry Erasmus Medical Center, Rotterdam, Netherlands.
226. Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands.
227. Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Scotland, UK.
228. Genetics of Complex Traits, University of Exeter Medical School, RILD Building University of Exeter, Exeter, EX2 5DW.
229. European Centre for Environment and Human Health, University of Exeter Medical School, The Knowledge Spa, Truro, TR1 3HD, UK.
230. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK.
231. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.
232. NIHR Cambridge Biomedical Research Centre, Level 4, Institute of Metabolic Science Box 289 Addenbrooke's Hospital Cambridge CB2 OQQ, UK.
233. University of Cambridge Metabolic Research Laboratories, Level 4, Institute of Metabolic Science Box 289 Addenbrooke's Hospital Cambridge CB2 OQQ, UK.
234. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia.
235. Department of Genetics, Harvard Medical School, Boston MA 02115 USA.

236. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115 USA.
237. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115 USA.
238. Division of Population Health Sciences and Education, St George's, University of London, London, SW17 0RE, UK.
239. Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA).
240. Center for Medical Systems Biology, Leiden, The Netherlands.
241. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, 85764, Germany.
242. Department of Statistics, University of Oxford, Oxford, UK.
243. Mount Sinai School of Medicine, New York, NY 10029, USA.
244. The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
245. Department of Preventive Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.

Author Contributions

AEJ, TWW, MFF, MGr, VAF, KY, and LB contributed equally to this manuscript. LAC, KEN, IBB, TOK, RJFL, and CTL supervised this project together. LAC, KEN, IBB, TOK, TWW, RJFL and AEJ conceived and

designed the study. LAC, RJFL, AEJ and TOK coordinated the collection of genome-wide association and interaction study results from the participating studies. The association and interaction results were contributed by SWvdL, MASi, SH, GJdB, HMDR and GP (AtheroExpress); AVS, TBH, GE, LJL, and VG (AGES study); KEN, MGr, AEJ, KY, EBoe, and PGL (ARIC study); JBW, NGM, RPSM, PAFM, ACH and GWM (AUSTWIN study); DPS (B58C study); GC, LJP, JoH, AWM, ALJ, and JBe (BHS study); CSch, TA, EBot and RJFL (BioMe); TT, DHe, LF (BLSA); BM, TMB, KDT, SC and BMP (CHS); YW, NRL, LSA and KLM (CLHNS study); ZK, PMV, TC, SBe, GWa, and PV (COLAUS study); JMart, IR, and CH (Croatia-Korcula study); VV, IK, and OPo (Croatia-Vis study); LY, AB, DT, SLo, BB and PF (DESIR study); RRau, TAL, PK, MHa, KSa, and RM (DR's EXTRA study); KF, NP, TE, and AMe (EGCUT study); JL, RAS, CL, and NJW (Ely study); CL, JL, RAS and NJW (EPIC), JHZ, RL, RAS, and NJW (EPIC-Norfolk study); NA, MCZ and CMvD (ERF study); IBB, MFF, JC and LB (Family Heart Study); JL, RAS, CL, RJFL and NJW (Fenland study); FX, JW, JSN, VAF, NLHC, CTL, CSF, and LAC (FramHS); MBo, FSC, KLM, and RNB (FUSION study); JT, LK, CSa, and HAK (FUSION2 study); MGo, BKK and CAB (Gendian); DJP, JEH, LJH, SP, CH and BHS (Generation Scotland); LFB, SLRK, MAJ, and PAP (GENOA study); SAh, FR, IB, GHa and PWF (GLACIER study); JE, CO, JOJ, MLor, AE and LV (GOOD study); TSA, THa and TIAS (GOYA study); BOT, CAM, SLV, TF, JNH and RSC (GxE); MHol, MNH, CP, AL and HVe (Health06 study); YJS, TRi, TRa, MASa, DCR and CB (HERITAGE Family Study); JASm, JDF, SLRK, WZhao, and DRW (HRS study); AUJ, KK, OLH, LLB, AJW, and KH (HUNT2 study); MC, DBr, SLu, NGl, JASt and DC (HYPERGENES); RJS, BS, KG, UdF, AH, ET and DBa (IMPROVE); TT, DHe, and SBa (InCHIANTI study); JMJ, MEJ, NGr, and OPe (Inter99 study); TWW, IMH, MEZ, MMN, MO, ALD, HG, MW, RRaw, BT, AP and KSt (KORA S3 and S4 studies); JVvVO, JMV, SSch, MASw and BHRW (Lifelines); WZhang, MLo, UA, SAf, JCC, and JSK (LOLIPOP study); MEK, GED, TBG, GS, JiH, and WM (LURIC study); UL, CAH, LLeM, SBu, and LH (MEC study); AMan, LJRT and YdIC (MESA study); MLa, JK, AJS, HMS, PSC, and NN (METSIM study); MKa, DM and CO (MrOS); MHor, MRJ and MIM (NFBC66 study); LQ, THu, QQ and DJH (NHS study); DK, KKO, JL, and AW (NSHD study); JMV, GWi, GL, JJH, EJCdG and DIB (NTR study); PN, AFW,

NDH, SW, HC and JFW (ORCADES study); AMah, CML, EI, LL and APM (PIVUS); NV, SJLB and PvdH (Prevend); ST, DJS, BMB, AJMdC, IF, RW, PES, NS and JWJ (PROSPER); LP, MCV, JEC, JBl and CB (QFS study); ND, MCZ, AGU and HT (RS1/RS2/RS3 study); JBr, SSa, DS, GRA, and FC (SardiNIA study); RJS, BS, BG, KL, AH and UdF (SCARFSHEEP); AT, RB, GHo, MN, HVö and HJG (SHIP study); BOT, CAM, SLV, TF, JNH and RSC (SPT); SK, GK, GD and PD (THISEAS); PJvdM, IMN, HS, AJO, CAH and MBr (TRAILS study); MMa, CM and TDS (TwinsUK study); AYC, LMR, PMR and DIC (WGHS study); NZ, SR, JG, CK and UP (WHI study); MKu, CL, JL, and MKi (Whitehall study); LPL, NHK, MJ, MKä, OTR and TL (YFS study). TWW, MGr, KY, JC, DHa, JSN, TSA, NLHC, FR, LX, QQ, JW and AEJ cleaned and quality checked the association and interaction results from the participating studies. TWW, KY, VAF, XD, JC, DHa, JSN, TSA, NLHC, LX and AEJ performed the meta-analyses. AYC, AEJ, LLB, MFF, TOK and LAC collected the supplementary information from the participating studies. AEJ, MGr, MFF, KY and VAF organized the supplementary tables. DHa, TWW, and AEJ provided look-up information from the GWAS meta-analysis of BMI, WAISTadjBMI and WHRadjBMI. JMarc provided lookup information from Smoking GWAS meta-analysis. AEJ performed the look-up in the NHGRI-EBI GWAS Catalog. MGr, XD, AEJ and ZK performed the analyses for variance explained by common variants in the SMK and nonSMK groups. MFF, KY, CTL, XD, LB and AEJ reviewed the literature for the identified loci. AEJ, KY, VAF and MG performed approximate conditional analyses. TWW conducted power and type 1 error simulations. AEJ produced heatmap and forest plots. JDE and ADJ carried out the lookups for Expression Quantitative Trait loci. JP, EL and CTL conducted eQTL analyses in the Framingham Heart Study. JTy and TFr conducted validation analyses in UKBB. AEJ, MGr, and KY conducted meta-analyses of GIANT and UKBB results. AEJ, TWW, MFF, MGr, KY, VAF, XD, LB, JMarc, TOK, CTL, JSN, RJFL, KEN and LAC wrote the manuscript.

Corresponding Authors:

Anne E. Justice

Department of Epidemiology, Gillings School of Global Public Health

University of North Carolina at Chapel Hill

Chapel Hill, NC, 27599

USA.

anne.justice@unc.edu

L. Adrienne Cupples

Department of Biostatistics

Boston University School of Public Health

801 Massachusetts Avenue

Boston, MA 02118

USA

adrienne@bu.edu

ABSTRACT

Few genome-wide association studies (GWAS) account for environmental exposures, like smoking, potentially impacting the overall trait variance when investigating the genetic contribution to obesity-related traits. Here, we use GWAS data from 51,080 current smokers and 190,178 nonsmokers (87% European descent) to identify loci influencing BMI and central adiposity, measured as waist circumference and waist-to-hip ratio both adjusted for BMI. We identify 23 novel genetic loci, and 9 loci with convincing evidence of gene-smoking interaction (GxSMK) on obesity-related traits. We show consistent direction of effect for all identified loci and significance for 18 novel and for 5 interaction loci in an independent study sample. These loci highlight novel biological functions, including response to oxidative stress, addictive behavior, and regulatory functions emphasizing the importance of accounting for environment in genetic analyses. Our results suggest that tobacco smoking may alter the genetic susceptibility to overall adiposity and body fat distribution.

INTRODUCTION

Recent genome-wide association studies (GWAS) have described loci implicated in obesity, body mass index (BMI), and central adiposity. Yet most studies have ignored environmental exposures with possibly large impacts on the trait variance[1, 2]. Variants that exert genetic effects on obesity through interactions with environmental exposures often remain undiscovered due to heterogeneous main effects and stringent significance thresholds. Thus, studies may miss genetic variants that have effects in subgroups of the population, such as smokers[3].

It is often noted that currently-smoking individuals display lower weight/BMI and higher waist circumference (WC) as compared to nonsmokers[4, 5, 6, 7, 8]. Smokers also have the smallest fluctuations in weight over approximately 20 years compared to those who have never smoked or have stopped smoking [9, 10]. Also, heavy smokers (>20 cigarettes per day [CPD]) and those that have smoked for more than 20 years are at greater risk for obesity than non-smokers or light to moderate smokers (<20 CPD)[11, 12]. Men and women gain weight rapidly after smoking cessation, suggesting that many people intentionally smoke for weight management[13]. It remains unclear why smoking cessation leads to weight gain or why long-term smokers maintain weight throughout adulthood, although studies suggest that tobacco use suppresses appetite[14, 15] or alternatively, smoking may result in an increased metabolic rate[14, 15]. Identifying genes that influence adiposity and interact with smoking may help us clarify pathways through which smoking influences weight and central adiposity[15].

A comprehensive study that evaluates smoking in conjunction with genetic contributions is warranted. Using GWAS data from the Genetic Investigation of Anthropometric Traits (GIANT) Consortium, we identified 23 novel genetic loci, and 9 loci with convincing evidence of gene-smoking interaction

(GxSMK) on obesity, assessed by BMI, and central obesity independent of overall body size, assessed by WC adjusted for BMI (WCadjBMI) and waist-to-hip ratio adjusted for BMI (WHRadjBMI). By accounting for smoking status, we focus both on genetic variants observed through their main effects and GxSMK effects to increase our understanding of their action on adiposity-related traits. These loci highlight novel biological functions, including response to oxidative stress, addictive behavior, and regulatory functions emphasizing the importance of accounting for environment in genetic analyses. Our results suggest that smoking may alter the genetic susceptibility to overall adiposity and body fat distribution.

RESULTS

GWAS discovery overview

We meta-analyzed study-specific association results from 57 Hapmap-imputed GWAS and 22 studies with Metabochip, including up to 241,258 (87% European descent) individuals (51,080 current smokers and 190,178 nonsmokers) while accounting for current smoking (SMK) (**Methods, Supplementary Fig. 1, Supplementary Tables 1-4**). For primary analyses, we conducted meta-analyses across ancestries and sexes. For secondary analyses, we conducted meta-analyses in European-descent studies alone and sex-specific meta-analyses (**Tables 1-4**, **Supplementary Data 1-6**). We considered four analytical approaches to evaluate the effects of smoking on genetic associations with adiposity traits (**Figure 1, Methods**). Approach 1 (SNPadjSMK) examined genetic associations after adjusting for SMK. Approach 2 (SNPjoint) considered the joint impact of main effects adjusted for SMK + interaction effects[16]. Approach 3 focused on interaction effects (SNPint); Approach 4 followed up loci from Approach 1 for interaction effects (SNPscreen). Results from Approaches 1-3 were considered genome-wide significant (GWS) with a P-value$<5\times10^{-8}$ while Approach 4 used Bonferroni adjustment after screening. Lead variants >500 kb from previous associations with BMI, WCadjBMI, and WHRadjBMI were considered novel. All association

results are reported with effect estimates oriented on the trait increasing allele in the current smoking stratum.

Across the three adiposity traits, we identified 23 novel associated genetic loci (6 for BMI, 11 for WCadjBMI, 6 for WHRadjBMI) and nine having significant GxSMK interaction effects (2 for BMI, 2 for WCadjBMI, 5 for WHRadjBMI) (**Figure 1, Tables 1-4**, **Supplementary Data 1-6**). We provide a comprehensive comparison with previously-identified loci[1, 2] by trait in supplementary material (**Supplementary Data 7, Supplementary Note 1**).

Accounting for Smoking Status

For primary meta-analyses of BMI (combined ancestries and sexes), 58 loci reached GWS in Approach 1 (SNPadjSMK) (**Supplementary Data 1, Supplementary Fig. 2-3**), including two novel loci near *SOX11*, and *SRRM1P2* (**Table 1**). Three more BMI loci were identified using Approach 2 (SNPjoint), including a novel locus near *CCDC93* (**Supplementary Fig. 4-5**). For WCadjBMI, 62 loci reached GWS for Approach 1 (SNPadjSMK) and two more for Approach 2 (SNPjoint), including eight novel loci near *KIF1B*, *HDLBP*, *DOCK3*, *ADAMTS3*, *CDK6*, *GSDMC*, *TMEM38B*, and *ARFGEF2* (**Table 1**, **Supplementary Data 2**, **Supplementary Fig. 2-5**). Lead variants near *PSMB10* from Approaches 1 and 2 (rs14178 and rs113090, respectively) are >500 kb from a previously-identified WCadjBMI-associated variant (rs16957304); however, after conditioning on the known variant, our signal is attenuated ($P=3.02\times10^{-2}$ and $P=5.22\times10^{-3}$), indicating that this finding is not novel. For WHRadjBMI, 32 loci were identified in Approach 1 (SNPadjSMK), including one novel locus near *HLA-C*, with no additional loci in Approach 2 (SNPjoint) (**Table 1**, **Supplementary Data 3, Supplementary Fig. 2-5**).

607 We used GCTA[17] to identify loci from our primary meta-analyses that harbor multiple independent SNPs
608 (**Methods, Supplementary Tables 5-7**). Conditional analyses revealed no secondary signals within 500
609 kb of our novel lead SNPs. Additionally, we performed conditional association analyses to determine if
610 our novel variants were independent of previous GWAS loci within 500 kb that are associated with
611 related traits of interest. All BMI-associated SNPs were independent of previously-identified GWS
612 associations with anthropometric and obesity-related traits. Seven novel loci for WCadjBMI were near
613 previous associations with related anthropometric traits. Of these, association signals for rs6743226
614 near *HDLBP*, rs10269774 near *CDK6*, and rs6012558 near *ARFGEF2* were attenuated (P>1E-5 and β
615 decreased by half) after conditioning on at least one nearby height and hip circumference adjusted for
616 BMI (HIPadjBMI) SNP, but association signals remained independent of other related SNP-trait
617 associations. For WHRadjBMI, our GWAS signal was attenuated by conditioning on two known height
618 variants (rs6457374 and rs2247056), but remained significant in other conditional analyses. Given high
619 correlations among waist, hip, and height, these results are not surprising.
620
621 Several additional loci were identified for Approaches 1 and 2 in secondary meta-analysis (**Table 2**,
622 **Supplementary Data 1-6**, **Supplementary Fig. 6**). For BMI, 2 novel loci were identified by Approach 1,
623 including 1 near *EPHA3* and 1 near *INADL*. For WCadjBMI, 2 novel loci were identified near *RAI14* and
624 *PRNP*. For WHRadjBMI, five novel loci were identified in secondary meta-analyses near *BBX, TRBI1*,
625 *EHMT2*, *SMIM2* and *EYA4*. A comprehensive summary of nearby genes for all novel loci and their
626 potential biological relevance is available in **Supplementary Note 2**.
627
628 **Figure 3** presents analytical power for Approaches 1 and 2 while **Supplementary Table 8** and
629 **Supplementary Fig. 7** present simulation results to evaluate type 1 error (**Methods)**. A heat map cross-
630 tabulates P-values for Approaches 1 and 2 along with Approach 3 examining interaction only

(**Supplementary Fig 8**). We demonstrate that the two approaches yield valid type 1 error rates and that Approach 1 can be more powerful to find associations given zero or negligible quantitative interactions, whereas Approach 2 is more efficient in finding associations when interaction exists.

Modification of Genetic Predisposition by Smoking

Approach 3 directly evaluated GxSMK interaction (SNPint) (**Table 3, Supplementary Data 1-6**, **Figure 2**, **Supplementary Fig. 9-10)**. For primary meta-analysis of BMI, two loci reached GWS including a previously identified GxSMK interaction locus near *CHRNB4*[3], and a novel locus near *INPP4B*. Both loci exhibit GWS effects on BMI in smokers and no effects in nonsmokers. For *CHRNB4* (cholinergic nicotine receptor B4), the variant minor allele (G) exhibits a decreasing effect on BMI in current smokers (βsmk = - 0.047) but no effect in nonsmokers (βnonsmk = 0.002). Previous studies identified nearby SNPs in high LD associated with smoking (nonsynonymous, rs16969968 in *CHRNA5*)[3] and arterial calcification (rs3825807, a missense variant in *ADAMTS7)*[18]. Conditioning on these variants attenuated our interaction effect but did not eliminate it (**Supplementary Table 7**), suggesting a complex relationship between smoking, obesity, heart disease, and genetic variants in this region. Importantly, the *CHRNA5-CHRNA3-CHRNB4* gene cluster has been associated with lower BMI in current smokers[3], but with higher BMI in never smokers[3], evidence supporting the lack of association in nonsmokers as well as a lack of previous GWAS findings on 15q25 (**Supplementary Data 8**)[1]. The *CHRNA5-CHRNA3-CHRNB4* genes encode the nicotinic acetylcholine receptor (nAChR) subunits α3, α5 and β4, which are expressed in the central nervous system[19]. Nicotine has differing effects on the body and brain, causing changes in metabolism and feeding behaviors[20]. These findings suggest smoking exposure may modify genetic effects on 15q24-25 to influence smoking-related diseases, such as obesity, through distinct pathways.

In primary meta-analyses of WCadjBMI, one novel GWS locus (near *GRIN2A*) with opposite effect directions by smoking status was identified for Approach 3 (SNPint) (**Table 3, Supplementary Data 2, Figure 2, Supplementary Fig. 9-10**). The T allele of rs4141488 increases WCadjBMI in current smokers and decreases it in nonsmokers (βsmk = 0.037, βnonsmk = -0.015). In secondary meta-analysis of European women-only, we identified an interaction between rs6076699, near *PRNP*, and SMK on WCadjBMI (**Table 4**, **Supplementary Data 5**, **Supplementary Fig. 6**), a locus also identified in Approach 2 (SNPjoint) for European women. The major allele, A, has a positive effect on current smokers as compared to a weaker and negative effect on WC in nonsmokers (βsmk = 0.169, βnonsmk = -0.070), suggesting why this variant remained undetected in previous GWAS of WCadjBMI (**Supplementary Data 8**).

Approach 4 (SNPscreen) (**Figure 1**, **Methods**) evaluated GxSMK interactions after screening SNPadjSMK results (from Approach 1) using Bonferroni-correction (**Methods**, **Tables 3-4**, **Supplementary Data 1-6**). We identified two SNPs, near *LYPLAL1* and *RSPO3*, with significant interaction; both have previously published main effects on anthropometric traits. These loci exhibit effects on WHRadjBMI in nonsmokers, but not in smokers (**Figure 2**). In secondary meta-analyses, we identified three known loci with significant GxSMK interaction effects on WHRadjBMI near *MAP3K1*, *HOXC4-HOXC6*, and *JUND* (**Table 4**, **Supplementary Data 3 and 6**). We identified rs1809420, near *CHRNA5-CHRNA3-CHRNB4*, for BMI in the men-only, combined-ancestries meta-analysis (**Supplementary Data 1**).

Power calculations demonstrate that Approach 4 has increased power to identify SNPs that show (i) an effect in one stratum (smokers or nonsmokers) and a less pronounced but concordant effect in the other stratum, or (ii) an effect in the larger nonsmoker stratum and no effect in smokers (**Figure 3**). In contrast, Approach 3 has increased power for SNPs that show (i) an effect in the smaller smoker stratum

and no effect in nonsmokers, or (ii) an opposite effect between smokers and nonsmokers (**Figure 3**). Our findings for both approaches agree with these power predictions, supporting usingboth analytical approaches to identify GxSMK interactions.

Enrichment of Genetic Effects by Smoking Status

When examining the smoking specific effects for BMI and WCadjBMI loci in our meta-analyses, no significant enrichment of genetic effects by smoking status were noted. (**Figure 2**, **Supplementary Fig. 11-12**). However, our results for WHRadjBMI were enriched for loci with a stronger effect in nonsmokers as compared to smokers, with 35 of 45 loci displaying numerically larger effects in nonsmokers ($P_{binomial}=1.2x10^{-4}$).

We calculated the variance explained by subsets of SNPs selected on 15 significance thresholds for Approach 1 from $P_{SNPadjSMK}=1x10^{-8}$ to $P_{SNPadjSMK}=0.1$ (**Supplementary Table 9**, **Figure 4**). Differences in variance explained between smokers and nonsmokers were significant (P<0.003=0.05/15, Bonferroni-corrected for 15 thresholds) for BMI at each threshold, with more variance explained in smokers. For WCadjBMI, the difference was significant for SNP sets beginning with $P_{SNPadjSMK}<3.16x10^{-4}$, and for WHRadjBMI at $P_{SNPadjSMK}<1x10^{-6}$. In contrast to BMI, SNPs from Approach 1 explained a greater proportion of the variance in nonsmokers for WHRadjBMI. Differences in variance explained were greatest for BMI (differences ranged from 1.8% - 21% for smokers) and lowest for WHRadjBMI (ranging from 0.3% to 8.8% for nonsmokers).

These results suggest that smoking may increase genetic susceptibility to overall adiposity, but attenuate genetic effects on body fat distribution. This contrast is concordant with phenotypic observations of higher overall adiposity and lower central adiposity in smokers[4, 5, 7, 8, 9]. Additionally,

smoking increases oxidative stress and general inflammation in the body[21] and may exacerbate weight gain[22]. Many genes implicated in BMI are involved in appetite regulation and feeding behavior[1]. For waist traits, our results adjusted for BMI likely highlight distinct pathways through which smoking alters genetic susceptibility to body fat distribution. Overall, our results indicate that more loci remain to be discovered as more variance in the trait can be explained as we drop the threshold for significance.

Functional or Biological Role of Novel Loci

We conducted thorough searches of the literature and publicly available bioinformatics databases to understand the functional role of all genes within 500 kb of our lead SNPs. We systematically explored the potential role of our novel loci in affecting gene expression both with and without accounting for the influence of smoking behavior (**Methods**, **Supplementary Note 3, Supplementary Tables 10-12**).

We found the majority of novel loci are near strong candidate genes with biological functions similar to previously identified adiposity-related loci, including regulation of body fat/weight, angiogenesis/adipogenesis, glucose and lipid homeostasis, general growth and development. (**Supplementary Notes 2** and **3**).

We identified rs17396340 for WCadjBMI (Approaches 1 and 2), an intronic variant in the *KIF1B* gene. This variant is associated with expression of *KIF1B* in whole blood with and without accounting for SMK (GTeX and **Supplementary Tables 10** and **12**) and is highly expressed in the brain[23]. Knockout and mutant forms of *KIF1B* in mice resulted in multiple brain abnormalities, including hippocampus morphology[24], a region involved in (food) memory and cognition[25]. Variant rs17396340 is associated with expression levels of *ARSA* in LCL tissue. Human adipocytes express functional *ARSA*, which turns

dopamine sulfate into active dopamine. Dopamine regulates appetite through leptin and adiponectin levels, suggesting a role for *ARSA* in regulating appetite[26].

Expression of *CD47* (CD47 molecule), near rs670752 for WHRadjBMI (Approach 1, women-only), is significantly decreased in obese individuals and negatively correlated with BMI, WC, and Hip circumference[27]. Conversely, in mouse models, CD47 deficient mice show decreased weight gain on high fat diets, increased energy expenditure, improved glucose profile, and decreased inflammation[28].

Several novel loci harbor genes involved in unique biological functions and pathways including addictive behaviors and response to oxidative stress. These potential candidate genes near our association signals are highly expressed in relevant tissues for regulation of adiposity and smoking behavior (e.g. brain, adipose tissue, liver, lung, muscle) (**Supplementary Note 2**, **Supplementary Table 10**).

The *CHRNA5-CHRNA3-CHRNB4* cluster is involved in the eNOS signaling pathway (Ingenuity KnowledgeBase, http://www.ingenuity.com) that is key for neutralizing reactive oxygen species introduced by tobacco smoke and obesity[29, 30]. Disruption of this pathway has been associated with dysregulation of adiponectin in adipocytes of obese mice, implicating this pathway in downstream effects on weight regulation[30, 31]. This finding is especially important due to the compounded stress adiposity places on the body as it increases chronic oxidative stress itself[31]. *INPP4B* has been implicated in the regulation of the PI3K/Akt signaling pathway[32] that is important for cellular growth and proliferation, but also eNOS signaling, carbohydrate metabolism, and angiogenesis[33].

GRIN2A, near rs4141488, controls long-term memory and learning through regulation and efficiency of synaptic transmission[34] and has been associated with heroin addiction[35]. Nicotine increases the

752 expression of *GRIN2A* in the prefrontal cortex in murine models[36]. There are no established relationships
753 between *GRIN2A* and obesity-related phenotypes in the literature, yet memantine and ketamine,
754 pharmacological antagonists of GRIN2A activity[37, 38, 39, 40], are implicated in treatment for obesity-
755 associated disorders, including binge-eating disorders and morbid obesity (ClinicalTrials.gov identifiers:
756 NCT00330655, NCT02334059, NCT01997515, NCT01724983). Memantine is under clinical investigation
757 for treatment of nicotine dependence (ClinicalTrials.gov identifiers: NCT01535040, NCT00136786,
758 NCT00136747). While our lead SNP is not within a characterized gene, rs4141488 and variants in high LD
759 (r^2>0.7) are within active enhancer regions for several tissues, including liver, fetal leg muscle, smooth
760 stomach and intestinal muscle, cortex, and several embryonic and pluripotent cell types
761 (**Supplementary Note 2**), and therefore may represent an important regulatory region for nearby genes
762 like *GRIN2A*.
763
764 In secondary meta-analysis of European women-only, we identified a significant GxSMK interaction for
765 rs6076699 on WCadjBMI (**Table 4**, **Supplementary Data 4**, **Supplementary Fig. 6**). This SNP is 100kb
766 upstream of *PRNP* (prion protein), a signaling transducer involved in multiple biological processes
767 related to the nervous system, immune system, and other cellular functions (**Supplementary Note 2**)[41].
768 Alternate forms of the oligomers may form in response to oxidative stress caused by copper exposure[42].
769 Copper is present in cigarette smoke and elevated in the serum of smokers, but is within safe ranges[43,
770 44]. Another gene near rs6076699, *SLC23A2* (Solute Carrier Family 23 [Ascorbic Acid Transporter],
771 Member 2), is essential for the uptake and transport of Vitamin C, an important nutrient for DNA and
772 cellular repair in response to oxidative stress both directly and through supporting the repair of Vitamin
773 E after exposure to oxidative agents[45, 46]. SLC23A2 is present in the adrenal glands and murine models
774 indicate that it plays an important role in regulating dopamine levels[47]. This region is associated with
775 success in smoking cessation and is implicated in addictive behaviors in general[48, 49]. Our tag SNP is

776 located within an active enhancer region (marked by open chromatin marks, DNAse hypersentivity, and
777 transcription factor binding motifs); this regulatory activity appears tissue specific (sex-specific tissues
778 and lungs) [HaploReg and UCSC Genome Browser].
779
780 Nicotinamide mononucleotide adenylyltransferease (*NMNAT1*), upstream of WCadjBMI variant
781 rs17396340, is responsible for the synthesis of NAD from ATP and NMN[50, 51]. NAD is necessary for
782 cellular repair following oxidative stress. Upregulation of *NMNAT* protects against damage caused by
783 reactive oxygen species in the brain, specifically the hippocampus[52, 53]. Also for WCadjBMI, both *CDK6*,
784 near SNP rs10269774, and *FAM49B*, near SNP rs6470765, are targets of the *BACH1* transcription factor,
785 involved in cellular response to oxidative stress and management of the cell cycle[54].
786
787 **Influence of Novel Loci on Related Traits**
788
789 In a look-up in existing GWAS of smoking behaviors (Ever/Never, Current/Not-Current, Smoking
790 Quantity [SQ])[55] (**Supplementary Data 8**), eight of our 26 SNPs were nominally associated with at least
791 one smoking trait. After multiple test correction ($P<0.05/26=0.0019$), only one SNP remains significant:
792 rs12902602, identified for Approaches 2 (SNPjoint) and 3 (SNPint) for BMI, showed association with SQ
793 ($P=1.45x10^{-9}$).
794
795 We conducted a search in the NHGRI-EBI GWAS Catalog[56, 57] to determine if any of our newly identified
796 loci are in high LD with variants associated with related cardiometabolic and behavioral traits or
797 diseases. Of the seven novel BMI SNPs, only rs12902602 was in high LD ($r^2>0.7$) with SNPs previously
798 associated with smoking-related traits (e.g. nicotine dependence), lung cancer, and cardiovascular
799 diseases (e.g. coronary heart disease) (**Supplementary Table 13**). Of the 12 novel WCadjBMI SNPs, five
800 were in high LD with previously-reported GWAS variants for mean platelet volume, height, infant length,

and melanoma. Of the six novel WHRadjBMI SNPs, three were near several previously associated variants, including cardiometabolic traits (e.g. LDL cholesterol, triglycerides, and measures of renal function).

Given high phenotypic correlation between WC and WHR with height, and established shared genetic associations that overlap our adiposity traits and height[1, 2, 58] we expect cross-trait associations between our novel loci and height. Therefore, we conducted a look-up of all of our novel SNPs to identify overlapping association signals (**Supplementary Data 8**). No novel BMI loci were significantly associated with height (P<0.002[0.05/24] SNPs). However, there are additional variants that may be associated with height, but not previously reported in GWAS examining height, including 2 for WHRadjBMI near *EYA4* and *TRIB1*, and 2 for WCadjBMI near *KIF1B* and *HDLBP* (P<0.002).

Lastly, as smoking has a negative (weight decreasing) effect on BMI, it is likely that smoking associated genetic variants have an effect on BMI in current smokers. Therefore, we expected that smoking associated SNPs exhibit some interaction with smoking on BMI. We looked up published smoking behavior SNPs[56, 57], 10 variants in 6 loci, in our own results. Two variants reached nominal significance (P<0.05) for GxSMK interaction on BMI (**Supplementary Table 14**), but only one reached Bonferroni-corrected significance (P<0.005). No smoking-associated SNPs exhibited GxSMK interaction. Therefore, we did not see a strong enrichment for low interaction P-values among previously identified smoking loci.

Validation of Novel Loci

We pursued validation of our novel and interaction SNPs in an independent study sample of up to 119,644 European adults from the UK Biobank study (**Tables 1-4**, **Supplementary Table 15,**

826 **Supplementary Fig 9**). We found consistent directions of effects in smoking strata (for Approaches 2 and
827 3) and in SNPadjSMK results (Approach 1) for each locus examined (**Supplementary Fig. 13**). For BMI, 3
828 SNPs were not GWS (P>5E-8) following meta-analysis with our GIANT results: rs12629427 near *EPAH3*
829 (Approach 1); rs1809420 within a known locus near *ADAMTS7* i(Approach 4) remained significant for
830 interaction, but not for SNPadjSMK; and rs336396 near *INPP4B* (Approach 3). For WCadjBMI, 3 SNPs
831 were not GWS (P>5E-8) following meta-analysis with our results: rs1545348 near *RAI14* (Approach 1);
832 rs4141488 near *GRIN2A* (Approach 3); and rs6012558 near *PRNP* (Approach 3). For WHRadjBMI, only 1
833 SNP from Approach 4 was not significant following meta-analysis with our results: rs12608504 near
834 *JUND* remained GWS for SNPadjSMK, but was only nominally significant for interaction (P_{int}=0.013).
835
836 **Challenges in Accounting for Environmental Exposures in GWAS**
837
838 A possible limitation of our study may be the definition and harmonization of smoking status. We chose
839 to stratify on current smoking status without consideration of type of smoking (e.g. cigarette, pipe) for
840 two reasons. First, focusing on weight alone, former smokers tend to return to their expected weight
841 quickly following smoking cessation[9, 15, 59]. Second, this definition allowed us to maximize sample size, as
842 many participating studies only had current smoking status available. However, WC and WHR may not
843 behave in the same manner as weight and BMI with former smokers retaining excess fat around their
844 waist[60]. Thus, results may differ with alternative harmonization of smoking exposure.
845
846 Another limitation may be potential bias in our effect estimates when adjusting for a correlated
847 covariate (e.g. collider bias)[61, 62]. This phenomenon is of particular concern when the correlation
848 between the outcome and the covariate is high and when significant genetic associations occur with
849 both traits in opposite directions. IOur analyses adjusted both WC and WHR for BMI. WHR has a
850 correlation of 0.49 with BMI, while WC has a correlation of 0.85[62]. Using previously published results for

BMI, WCadjBMI and WHRadjBMI, we find three novel loci for WCadjBMI (near *DOCK3, ARFGEF2, TMEM38B*) and two for WHRadjBMI (near *EHMT2, HLA-C*) (**Supplementary Data 8**) with nominally significant associations with BMI and opposite directions of effect. At these loci, the genetic effect estimates should be interpreted with caution. Additionally, we adjusted for SMK in Approach 1 (SNPadjSMK). However binary smoking status, as we used, has a low correlation to BMI, WC, and WHR, as estimated in the ARIC study's European descent participants (-0.13, 0.08, and 0.12 respectively) and in the Framingham Heart Study (-0.05, 0.08, 0.16). Additionally, there are no loci identified in Approach 1 (SNPadjSMK) that are associated with any smoking behavior trait and that exhibit an opposite direction of effect from that identified in our adiposity traits (**Supplementary Data 8**). We therefore preclude potential collider bias and postulate true gain in power through SMK-adjustment at these loci.

To assess how much additional information is provided by accounting for SMK and GxSMK in GWAS for obesity traits, we compared genetic risk scores (GRSs) based on various subsets of lead SNP genotypes in various regression models (**Methods**). While any GRS was associated with its obesity trait ($P<1.6 \times 10^{-7}$, **Supplementary Table 16**), adding SMK and GxSMK terms to the regression model along with novel variants to the GRSs substantially increased variance explained. For example, variance explained increased by 38% for BMI (from 1.53% to 2.11%, $P=4.3\times10^{-5}$), by 27% for WCadjBMI (from 2.59% to 3.29%, $P=3.9\times10^{-6}$) and by 168% for WHRadjBMI (from 0.82% to 2.20%, $P=3.2\times10^{-11}$). Therefore, despite potential limitations, there is much to be gained by accounting for environmental exposures in GWAS studies.

DISCUSSION

To better understand the effects of smoking on genetic susceptibility to obesity, we conducted meta-analyses to uncover genetic variants that may be masked when the environmental influence of smoking is not considered, and to discover genetic loci that interact with smoking on adiposity-related traits. We identified 161 loci in total, including 23 novel loci (6 for BMI, 11 for WCadjBMI, and 6 for WHRadjBMI). While many of our newly identified loci support the hypothesis that smoking may influence weight fluctuations through appetite regulation, these novel loci also have highlighted new biological processes and pathways implicated in the pathogenesis of obesity.

Importantly, we identified nine loci with convincing evidence of GxSMK interaction on obesity-related traits. We were able to replicate the previous GxSMK interaction with BMI within the *CHRNA5-CHRNA3-CHRNB4* gene cluster. One novel BMI-associated locus near *INPP4B* and two novel WCadjBMI-associated loci near *GRIN2A* and *PRNP* displayed significant GxSMK interaction. We were also able to identify significant GxSMK interaction for one known BMI-associated locus near *ADAMTS7* and for five known WHRadjBMI-associated loci near *LYPLAL1*, *RSPO3*, *MAP3K1*, *HOXC4-HOXC6* and *JUND*. The majority of these loci harbor strong candidate genes for adiposity with a possible role for the modulation of effects through tobacco use.

We identified 18 new loci in Approach 1 ($P_{SNPadjSMK}$) by adjusting for current smoking status. Our analyses did not allow us to determine whether these discoveries are due to different subsets of subjects included in the analyses compared to previous studies[1, 2] or due only to adjusting for current smoking. Adjustment for current smoking in our analyses, however, did reveal novel associations. Specifically after accounting for smoking in our analyses, all novel BMI loci exhibit P-values that are at least one order of magnitude lower than in previous GIANT investigations, despite smaller samples in the current analysis[2]. While sample sizes for both WCadjBMI and WHRadjBMI are comparable with previous GIANT

investigations, our p-values for variants identified in Approach 1 are at least two orders of magnitude lower than previous findings. Thus, adjustment for smoking may have indeed revealed new loci. Further, loci identified in Approach 2, including 9 novel loci, suggest that accounting for interaction improves our ability to detect these loci even in the presence of only modest evidence of GxSMK interaction.

There are several challenges in validating genetic associations that account for environmental exposure. In addition to exposure harmonization and potential bias due to adjustment for smoking exposure, differences in trait distribution, environmental exposure frequency, ancestry-specific LD patterns and allele frequency across studies may lead to difficulties in replication, especially for gene-by-environment studies[63, 64]. Further, the "winner's curse" (inflated discovery effects estimates) requires larger sample sizes for adequate power in replication studies[65]. Despite these challenges, we were able to detect consistent direction of effect in an independent sample for all novel loci. Some results that did not remain GWS in the GIANT + UKBB meta-analysis had results that were just under the threshold for significance, suggesting that a larger sample may be needed to confirm these results, and thus the associations near *INPP4B*, *GRIN2A*, *RAI14*, *PRNP*, and *JUND* should be interpreted with caution.

While we found that effects were not significantly enriched in smokers for BMI, there is a greater proportion of variance in BMI explained by variants that are significant for Approach 1 (SNPadjSMK), which may be expected given that there are a greater number of variants with higher effect estimates in smokers. For WCadjBMI, there was no enrichment for stronger effects in one stratum compared to the other for our significant loci; however, there was a greater proportion of explained variance in WCadjBMI for loci identified in Approach 1 (SNPadjSMK) in nonsmokers. For WHRadjBMI, there were significantly more loci that exhibit greater effects in nonsmokers, and this pattern was mirrored in the variance explained analysis. The large difference between effects in smokers and nonsmokers likely

explains the sub-GWS levels of our loci in previous GIANT investigations[2]. For example, the T allele of rs7697556, 81kb from the *ADAMTS3* gene, was associated with increased WCadjBMI and exhibits a six-fold greater effect in nonsmokers compared to smokers, although the interaction effect was only nominal; in previous GWAS this variant was nearly GWS. These differences in effect estimates between smokers and nonsmokers may help explain inconsistent findings in previous analyses that show central adiposity increases with increased smoking, but is associated with decreased weight and BMI[6, 11, 12].

Our results support previous findings that implicate genes involved in transcription and gene expression, appetite regulation, macronutrient metabolism, and glucose homeostasis. Several of our novel loci have candidate genes within 500 kb of our tag variants that are highly expressed and/or active in brain tissue (*BBX*, *KIF1B*, *SOX11*, and *EPHA3*) and, like other obesity-associated genes, may be involved in previously-identified pathways linked to neuronal regulation of appetite (*KIF1B*, *GRIN2A*, and *SLC23A2*), adipo/angiogenesis (*ANGPTL3* and *TNF*) and glucose, lipid and energy homeostasis (*CD47*, *STK25*, *STK19*, *RAGE*, *AIF1*, *LYPLAL1*, *HDLBP*, *ANGPTL3*, *DOCK7*, *KIF1B*, *PREX1*, and *RPS12*).

Many our newly identified loci highlight novel biological functions and pathways where dysregulation may lead to increased susceptibility to obesity, including response to oxidative stress, addictive behavior, and newly identified regulatory functions. There is a growing body of evidence that supports the notion that exposure to oxidative stress leads to increased adiposity, risk of obesity, and poor cardiometabolic outcomes[30, 66, 67]. Our results for BMI and WCadjBMI, specifically associations identified near *CHRNA5-CHRNA3-CHRNB4, PRNP, SLC23A2, BACH1, and NMNAT1,* highlight new biological pathways and processes for future examination and may lead to a greater understanding of how oxidative stress leads to changes in obesity phenotypes and downstream cardiometabolic risk.

By considering current smoking, we were able to identify 6 novel loci for BMI, 11 for WCadjBMI, and 6 for WHRadjBMI, and highlight novel biological processes and regulatory functions for genes implicated in increased obesity risk. Eighteen of these remained significant in our validation with the UK Biobank sample. We confirmed most established loci in our analyses after adjustment for smoking status in smaller samples than were needed in previous discovery analyses. A typical approach in large-scale GWAS meta-analyses is not to adjust for covariates such as current smoking; our findings highlight the importance of accounting for environmental exposures in genetic analyses.

METHODS

Study Design Overview

We applied four approaches to identify genetic loci that influence adiposity traits by accounting for current tobacco smoking status (**Figure 1**). We defined smokers as those who responded that they were currently smoking; not current smokers were those that responded "no" to currently smoking. We evaluated three traits: body mass index (BMI), waist circumference adjusted for BMI (WCadjBMI), and waist-to-hip ratio adjusted for BMI (WHRadjBMI). Our first two meta-analytical approaches were aimed at determining whether there are novel genetic variants that affect adiposity traits by adjusting for SMK (SNPadjSMK), or by jointly accounting for SMK and for interaction with SMK (SNPjoint); while Approaches 3 and 4 aimed to determine whether there are genetic variants that affect adiposity traits through interaction with SMK (SNPint and SNPscreen) (**Figure 1**). Our ***primary meta-analyses*** focused on results from all ancestries, sexes combined. ***Secondary meta-analyses*** were performed using the European-descent populations only, as well as stratified by sex (men-only and women-only) in all ancestries and in European-descent study populations.

Cohort Descriptions and Sample Sizes

The GIANT consortium was formed by an international group of researchers interested in understanding the genetic architecture of anthropometric traits (**Supplemental Tables 1-4** for study sample sizes and descriptive statistics). In total, we included up to 79 studies comprising up to 241,258 individuals for BMI (51,080 smokers, 190,178 nonsmokers), 208,176 for WCadjBMI (43,226 smokers, 164,950 nonsmokers), and 189,180 for WHRadjBMI (40,543 smokers, 148,637 nonsmokers) with HapMap II imputed genome-wide chip data (up to 2.8M SNPs in association analyses), and/or with genotyped MetaboChip data

(~195K SNPs in association analyses)[68]. In instances where studies submitted both Metabochip and GWAS data, these were for non-overlapping individuals. Each study's Institutional Review Board has approved this research and all study participants have provided written informed consent.

Phenotype descriptions

Our study highlights three traits of interest: BMI, WCadjBMI and WHRadjBMI. Height and weight, used to calculate BMI (kg/m^2), were measured in all studies; waist and hip circumferences were measured in the vast majority. For each sex, traits were adjusted using linear regression for age and age^2 (as well as for BMI for WCadjBMI and WHRadjBMI), and (when appropriate) for study site and principal components to account for ancestry. Family studies used linear mixed effects models to account for familial relationships and also conducted analyses for men and women combined including sex in the model. Phenotype residuals were obtained from the adjustment models and were inverse normally transformed subsequently to facilitate comparability across studies and with previously published analyses. The trait transformation was conducted separately for smokers and nonsmokers for the SMK-stratified model and using all individuals for the SMK-adjusted model.

Defining Smokers

The participating studies have varying levels of information on smoking, some with a simple binary variable and others with repeated, precise data. Since the effects of smoking cessation on adiposity appear to be immediate[9, 10, 59], a binary smoking trait (current smoker vs. not current smoker) is used for the analyses as most studies can readily derive this variable. We did not use a variable of 'ever smoker vs. never' as it increases heterogeneity across studies, thus adding noise; also this definition would make harmonization across studies difficult.

Genotype Identification and Imputation

Studies with GWAS array data or Metabochip array data contributed to the results. Each study applied study-specific standard exclusions for sample call rate, gender checks, sample heterogeneity and ethnic group outliers (**Supplementary Table 2**). For most studies (except those that employed directly typed MetaboChip genotypes), genome-wide chip data was imputed to the HapMap II reference data set via MACH[69], IMPUTE[70], BimBam[71] or Beagle[72].

Study Level Analyses

To obtain study-specific summary statistics used in subsequent meta-analyses, the following linear models (or linear mixed effects models for studies with families/related individuals) were run separately for men and women and separately for cases and controls for case-control studies using phenotype residuals from the models described above. Studies with family data also conducted analyses with these models for men and women combined after accounting for dependency among family members as a function of their kinship correlations. We assumed an additive genetic model.

SMK-adjusted: $TRAIT = \beta_0 + \beta_1 SNP + \beta_2 SMK$

SMK-stratified: $TRAIT = \beta_0 + \beta_1 SNP$ (run in current smokers and nonsmokers separately)

The analyses were run using various GWAS software, including MACH2QTL [73], SNPTEST [74], ProbABEL[75], GenABEL[76], Merlin[77], PLINK[78] or QUICKTEST[79].

Quality control of study-specific summary statistics

The aggregated summary statistics were quality-controlled according to a standardized protocol[80]. These included checks for issues with trait transformations, allele frequencies and strand. Low quality SNPs in

each study were excluded for the following criteria: (i) SNPs with low minor allele count (MAC <= 5, MAC = MAF * N) and monomorphic SNPs, (ii) genotyped SNPs with low SNP call-rate (< 95%) or low Hardy-Weinberg equilibrium test P-Value (< 10^{-6}), (iii) imputed SNPs with low imputation quality (MACH-Rsq or OEVAR<0.3, or information score <0.4 for SNPTEST/IMPUTE/IMPUTE2, or <0.8 for PLINK). To test for issues with relatedness or overlapping samples and to correct for potential population stratification, the study-specific standard errors and association P-Values were genomic control (GC) corrected using lambda factors[81] (**Supplementary Fig. 1**). GC correction for GWAS data used all SNPs, but GC correction for MetaboChip data were restricted to chip QT interval SNPs only as the chip was enriched for associations with obesity-related traits. Any study-level GWAS file with a lambda > 1.5 was removed from further analyses. While we established this criterion, no study results were removed for this reason.

Meta-analyses

Meta-analyses used study-specific summary statistics for the phenotype associations for each of the above models. We used a fixed-effects inverse variance weighted method for the SNP main effect analyses. All meta-analyses were run in METAL[82]. As study results came in two separate batches (Stage 1 and Stage 2), meta-analyses from the two stages were further meta-analyzed (Stage 1 + Stage 2). A second GC correction was applied to all SNPs when combining Stage 1 and Stage 2 meta-analyses in the final meta-analysis. First, Hapmap-imputed GWAS data were meta-analyzed together, as were Metabochip studies. This step was followed by a combined GWAS + Metabochip meta-analysis. For primary analyses, we conducted meta-analyses across ancestries and sexes. For secondary meta-analyses, we conducted meta-analyses in European-descent studies alone, and sex-specific meta-analyses. There were two reasons for conducting secondary meta-analyses. First, both WCadjBMI and WHRadjBMI have been shown to display sex-specific genetic effects[2, 83, 84]. Second, by including

populations from multiple ancestries in our primary meta-analyses, we may be introducing heterogeneity due to differences in effect sizes, allele frequencies, and patterns of linkage disequilibrium across ancestries, potentially decreasing power to detect genetic effects. See **Supplementary Fig. 1** for a summary of the primary meta-analysis study design. The obtained SMK-stratified summary statistics were later used to calculate summary SNPjoint and SNPint statistics using EasyStrata[85]. Briefly, this software implements a two-sample, large sample test of equal regression parameters between smokers and nonsmokers as described by Randall et al[83] for SNPint and the two degree of freedom test of main and interaction effects for SNPjoint as described by Aschard et al[16].

Lead SNP selection

Before selecting a lead SNP for each locus, SNPs with high heterogeneity $I^2 \geq 0.75$ or a minimum sample size below 50% of the maximum N for each strata (e.g. N> max[N Women Smokers]/2) were excluded. Lead SNPs that met significance criteria were selected based on distance (+/- 500 kb), and we defined the SNP with the lowest P-value as the top SNP for a locus. SNPs that reached genome-wide significance (GWS), but had no other SNPs within 500 kb with a P<1E-5 (lonely SNPs), were excluded from the SNP selection process. Two variants were excluded from Approach 2 based on this criterion, rs2149656 for WCadjBMI and rs2362267 for WHRadjBMI.

Approaches

Figure 1 outlines the four approaches that we used to identify novel SNPs. The left side of Figure 1 focuses on the first hypothesis that examines the effect of SNPs on adiposity traits. *Approach 1* considered a linear regression model that includes the SNP and SMK, thus adjusting for SMK (SNPadjSMK). Summary SNPadjSMK results were obtained from the SMK-adjusted meta-analysis. *Approach 2* used summary SMK-stratified meta-analysis results as described by Aschard et al.[16] to

consider the joint hypothesis that a genetic variant has main and/or interaction effects on outcomes as a 2 degree of freedom test (SNPjoint). For this approach, the null hypothesis was that there is no main and no interaction effect on the outcome. Thus, rejection of this hypothesis could be due to either a main effect or an interaction effect or to both.

The right side of Figure 1 focuses on our second hypothesis, testing for interaction of a variant with SMK on adiposity traits as outcomes. *Approach 3* used the SMK-stratified results to directly contrast the regression coefficients for a test of interaction (SNPint)[83]. *Approach 4* used a screening strategy to evaluate interaction, whereby the SMK-adjusted main effect results (Approach 1) were screened for variants significant at the $P<5x10^{-8}$ level. These variants were then carried forward for a test of interaction, comparing the SMK-stratified specific regression coefficients in the second step (SNPscreen).

In *Approaches 1-3* variants significant at $P<5x10^{-8}$ were considered GWS. In *Approach 4* (SNPscreen) variants for which the p-value of the test of interaction is less than 0.05 divided by the number of variants carried forward were considered significant for interaction. We performed analytical power computations to demonstrate the usefulness and characteristic of the two interaction Approaches.

LocusZoom Plots

Regional association plots were generated for novel loci using the program Locuszoom[86]. For each plot, LD was calculated using a multiethnic sample of the 1000 Genomes Phase I reference panels[87], including EUR, AFR, EAS, and AMR. Previous SNP-trait associations highlighted within the plots include traits of interest (e.g. cardiometabolic, addiction, behavior, anthropometrics) found in the NHGRI-EMI GWAS Catalog and supplemented with recent GWAS studies from the literature[1, 2, 58, 84].

Conditional Analyses

To determine if multiple association signals were present within a single locus, we used GCTA[17] to perform approximate joint conditional analyses on the SNPadjSMK and SMK- stratified data. The following criteria were used to select candidate loci for conditional analyses: nearby SNP (+/- 500kb) with an R^2>0.4 and an association P<1E-5 for any of our primary analyses. GCTA uses associations from our meta-analyses and LD estimates from reference data sets containing individual-level genotypic data to perform the conditional analyses. To calculate the LD structure, we used two U.S. cohorts, the Atherosclerosis Risk in Communities (ARIC) study consisting of 9,713 individuals of European descent and 580 individuals of African American descent, and the Framingham Heart Study (FramHS) consisting of 8,481 individuals of European ancestry, both studies imputed to HapMap r22. However, because our primary analyses were conducted in multiple ancestries, each study supplemented the genetic data using HapMap reference populations so that the final reference panel was composed of about 1-3% Asians (CHB + JPT) and 4-6% Africans (YRI for the FramHS) for the entire reference sample. We extracted each 1 MB region surrounding our candidate SNPs, performed joint approximate conditional analyses, and then repeated the steps for the appropriate Approach to identify additional association signals.

Many of the SNPs identified in the current analyses were nearby SNPs previously associated with related anthropometric and obesity traits (e.g. height, visceral adipose tissue). For all lead SNPs near a SNP previously associated with these traits, GCTA was also used to perform approximate conditional analyses on the SNPadjSMK and SMK-stratified data in order to determine if the loci identified here are independent of the previously identified SNP-trait associations.

Power and Type I Error

In order to illustrate the validity of the approaches with regards to type 1 error, we conducted simulations. For two MAF, we assumed standardized stratum-specific outcomes for 50,000 smokers and 180,000 nonsmokers and generated 10,000 simulated stratum-specific effect sizes under the stratum-specific null hypotheses of "no stratum-specific effects". We applied the four approaches to the simulated stratum-specific association results and inferred type 1 error of each approach by visually examining QQ plots and by calculating type 1 error rates. The type 1 error rates shown reflect the proportion of nominally significant simulation results for the respective approach. Analytical power calculations to identify effects for various combinations of SMK- and NonSMK-specific effects by the Approaches 1-4 again assumed 50,000 smokers and 180,000 nonsmokers. We first assumed three different fixed effect estimates in smokers that were small (R^2_{SMK}=0.01%, similar to the realistic *NUDT3* effect on BMI), medium (R^2_{SMK}=0.07%, similar to the realistic *BDNF* effect on BMI) or large (R^2_{SMK}=0.34%, similar to the realistic *FTO* effect on BMI) genetic effects, and varied the effect in nonsmokers. Second, we assumed fixed (small, medium and large) effects in nonsmokers and varied the effect in smokers.

Biological Summaries

To identify genes that may be implicated in the association between our lead SNPs (Tables 1-3) and BMI, WHRadjBMI, and WCadjBMI, and to shed light on the complex relationship between genetic variants, SMK and adiposity, we performed in-depth literature searches on nearby candidate genes. Snipper v1.2 (http://csg.sph.umich.edu/boehnke/snipper/) was used to identify any genes and cis- or trans-eQTLs within 500kb of our lead SNPs. All genes identified by Snipper were manually curated and examined for evidence of relationship with smoking and/or adiposity. To explore any potential regulatory or function role of the association regions, loci were also examined using several bioinformatic tools/databases, including HaploReg v4.1[88], UCSC Genome Browser[89] (available at http://genome.ucsc.edu/), GTeX Portal[90], and RegulomeDB[91].

eQTL Analyses

We used two approaches to systematically explore the role of novel loci in regulating gene expression. First, to gain a general overview of the regulatory role of newly identified GWAS regions, we conducted an eQTL lookup using >50 eQTL studies[92], with specific citations for >100 datasets included in the current query for blood cell related eQTL studies and relevant non-blood cell tissue eQTLs (e.g. adipose and brain tissues). Additional eQTL data was integrated from online sources including ScanDB, the Broad Institute GTEx Portal, and the Pritchard Lab (eqtl.uchicago.edu). Additional details on the methods, including study references can be found in **Supplementary Note 3**. Only significant cis-eQTLS in high LD with our novel lead SNPs (r2>0.9, calculated in the CEU+YRI+CHB+JPT 1000 Genomes reference panel), or proxy SNPs, were retained for consideration.

Second, since public databases with eQTL data do not have information available on current smoking status, we also conducted a cis-eQTL association analysis using expression results derived from fasting peripheral whole blood using the Human Exon 1.0 ST Array (Affymetrix, Inc., Santa Clara, CA). The raw expression data were quantile-normalized, log2 transformed, followed by summarization using Robust Multi-array Average[93] and further adjusted for technical covariates, including the first principal component of the expression data, batch effect, the all-probeset-mean residual, blood cell counts, and cohort membership. We evaluated all transcripts +/- 1MB around each novel variant in the Framingham Heart Study while accounting for current smoking status, using the following four approaches similar to those used in our primary analyses of our traits: 1) eQTL adjusted for SMK, 2) eQTL stratified by SMK, 3) eQTL x SMK interaction, and 4) joint main + eQTLxSMK interaction). Significance level was evaluated by FDR < 5% per eQTL analysis and across all loci identified for that model in the primary meta-analysis. Additional details can be found in **Supplementary Note 3**.

Variance-explained estimates

We estimated the phenotypic variance in smokers and nonsmokers explained by the association signals using a method previously described by Kutalik et al.[94] For each associated region, we selected subsets of SNPs within 500 kb of our lead SNPs and based on varying P value thresholds (ranging from $1x10^{-8}$ to 0.1) from Approach 1 (SNPadjSMK model). First, each subset of SNPs was clumped into independent regions to identify the lead SNP for each region. The variance explained by each subset of SNPs in the SMK and nonSMK strata was estimated by summing the variance explained by the individual lead SNPs.

Smoking Behavior Lookups

In order to determine if any of the loci identified in the current study are associated with smoking behavior, we conducted a look-up of all lead SNPs from novel loci and Approach 3 in existing GWAS of smoking behavior[3]. The analysis consists of phasing study-specific GWAS samples contributing to the smoking behavior meta-analysis, imputation, association testing and meta-analysis. To ensure that all SNPs of interest were available in the smoking GWAS, the program SHAPEIT2[95] was used to phase a region 500Kb either side of each lead SNP, and imputation was carried out using IMPUTE2[96] with the 1000 Genomes Phase 3 dataset as a reference panel.

Each region was analyzed for 3 smoking related phenotypes: (i) Ever vs Never smokers, (ii) Current vs Non-current smokers, and (iii) a categorical measure of smoking quantity[55]. The smoking quantity levels were 0 (defined as 1-10 cigarettes per day [CPD]), 1 (11-20 CPD), 2 (21-30 CPD) and 3 (31 or more CPD). Each increment represents an increase in smoking quantity of 10 cigarettes per day. There were 10,058 Never smokers, 13,418 Ever smokers, 11,796 Non-current smokers, 6,966 Current smokers and 11,436 samples with the SQ phenotypes. SNPMETA[55] was used to perform an inverse-variance weighted fixed

effects meta-analysis across cohorts at all SNPs in each region, and included a single GC correction. At each SNP, only those cohorts that had an imputation info score > 0.5 were included in the meta-analysis.

Main Effects Lookup in Previous GIANT Investigations

To better understand why our novel variants remained undiscovered in previous investigations that did not take SMK into account, we also conducted a lookup of our novel variants in published GWAS results examining genetic main effects on BMI, WC, WCadjBMI, WHR, WHRadjBMI, and height[1, 2, 58].

GWAS Catalog Lookups

To further investigate the identified genetic variants in this study and to gain additional insight into their functionality and possible effects on related cardiometabolic traits, we searched for previous SNP-trait associations nearby our lead SNPs. PLINK was used to find all SNPs within 500 kb of any of our lead SNPs and calculate r^2 values using a combined ancestry (AMR, AFR, EUR, ASN) 1000 Genomes Phase 1 reference panel[87] to allow for LD calculation for SNPs on the Illumina Metabochip and to best estimate LD in our multiethnic GWAS. All SNPs within the specified regions were compared with the NHGRI-EBI (National Human Genome Research Institute, European Bioinformatics Institute) GWAS Catalog, version 1.0 (www.ebi.ac.uk/gwas)[56, 57] for overlap, and distances between the two SNPs were calculated using STATA v14, for the chromosome and base pair positions based on human genome reference build 19. All previous associations within 500 kb and with an R^2>0.5 with our lead SNP were retained for further interrogation.

Genetic risk score calculation

We calculated several unweighted genetic risk scores (GRSs) for each individual in the population-based KORA-S3 and KORA-S4 studies (total N = 3,457). We compared GRSs limited to previously known lead

SNPs (see **Supplementary Data 7** for lists of previously known lead SNPs) with GRSs based on previously known and novel lead SNPs from the current study (see **Tables 1-4** for lists of novel lead SNPs). Risk scores were tested for association with the obesity trait using the following linear regression models: The unadjusted GRS model (TRAIT = $\beta_0 + \beta_1$GRS), the adjusted GRS model (TRAIT = $\beta_0 + \beta_1$GRS + β_2SMK) and the GRSxSMK interaction model (TRAIT = $\beta_0 + \beta_1$GRS + β_2SMK + β_3GRSxSMK).

DATA AVAILABILITY

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

REFERENCES

1. Locke AE*, et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).

2. Shungin D*, et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187-196 (2015).

3. Taylor AE*, et al.* Stratification by smoking status reveals an association of CHRNA5-A3-B4 genotype with body mass index in never smokers. *PLoS genetics* **10**, e1004799 (2014).

4. Pistelli F, Aquilini F, Carrozzi L. Weight gain after smoking cessation. *Monaldi archives for chest disease = Archivio Monaldi per le malattie del torace / Fondazione clinica del lavoro, IRCCS [and] Istituto di clinica tisiologica e malattie apparato respiratorio, Universita di Napoli, Secondo ateneo* **71**, 81-87 (2009).

5. Flegal KM, Troiano RP, Pamuk ER, Kuczmarski RJ, Campbell SM. The influence of smoking cessation on the prevalence of overweight in the United States. *The New England journal of medicine* **333**, 1165-1170 (1995).

6. Morris RW*, et al.* Heavier smoking may lead to a relative increase in waist circumference: evidence for a causal relationship from a Mendelian randomisation meta-analysis. The CARTA consortium. *BMJ open* **5**, e008808 (2015).

7. Barrett-Connor E, Khaw KT. Cigarette smoking and increased central adiposity. *Annals of internal medicine* **111**, 783-787 (1989).

8. Kim JH, Shim KW, Yoon YS, Lee SY, Kim SS, Oh SW. Cigarette smoking increases abdominal and visceral obesity but not overall fatness: an observational study. *PloS one* **7**, e45815 (2012).

9. Owen-Smith V, Hannaford PC. Stopping smoking and body weight in women living in the United Kingdom. *The British journal of general practice : the journal of the Royal College of General Practitioners* **49**, 989-990 (1999).

10. Lahmann PH, Lissner L, Gullberg B, Berglund G. Sociodemographic factors associated with long-term weight gain, current body fatness and central adiposity in Swedish women. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* **24**, 685-694 (2000).

11. Chiolero A, Jacot-Sadowski I, Faeh D, Paccaud F, Cornuz J. Association of cigarettes smoked daily with obesity in a general adult population. *Obesity (Silver Spring)* **15**, 1311-1318 (2007).

12. Clair C*, et al.* Dose-dependent positive association between cigarette smoking, abdominal obesity and body fat: cross-sectional data from a population-based survey. *BMC public health* **11**, 23 (2011).

13. Pirie PL, Murray DM, Luepker RV. Gender differences in cigarette smoking and quitting in a cohort of young adults. *American journal of public health* **81**, 324-327 (1991).

14. Nicklas BJ, Tomoyasu N, Muir J, Goldberg AP. Effects of cigarette smoking and its cessation on body weight and plasma leptin levels. *Metabolism: clinical and experimental* **48**, 804-808 (1999).

15. Munafo MR, Johnstone EC, Murphy MF, Aveyard P. Lack of association of DRD2 rs1800497 (Taq1A) polymorphism with smoking cessation in a nicotine replacement therapy randomized trial. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco* **11**, 404-407 (2009).

16. Aschard H, Hancock DB, London SJ, Kraft P. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Human heredity* **70**, 292-300 (2010).

17. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82 (2011).

18. van Setten J*, et al.* Genome-wide association study of coronary and aortic calcification implicates risk loci for coronary artery disease and myocardial infarction. *Atherosclerosis* **228**, 400-405 (2013).

19. Antolin-Fontes B, Ables JL, Gorlich A, Ibanez-Tallon I. The habenulo-interpeduncular pathway in nicotine aversion and withdrawal. *Neuropharmacology* **96**, 213-222 (2015).

20. Picciotto MR, Mineur YS. Molecules and circuits involved in nicotine addiction: The many faces of smoking. *Neuropharmacology* **76 Pt B**, 545-553 (2014).

21. van der Vaart H, Postma DS, Timens W, ten Hacken NH. Acute effects of cigarette smoke on inflammation and oxidative stress: a review. *Thorax* **59**, 713-721 (2004).

22. Aroor AR, DeMarco VG. Oxidative stress and obesity: the chicken or the egg? *Diabetes* **63**, 2216-2218 (2014).

23. Yang HW, Chen YZ, Takita J, Soeda E, Piao HY, Hayashi Y. Genomic structure and mutational analysis of the human KIF1B gene which is homozygously deleted in neuroblastoma at chromosome 1p36.2. *Oncogene* **20**, 5075-5083 (2001).

24. Zhao C*, et al.* Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1Bbeta. *Cell* **105**, 587-597 (2001).

25. Davidson TL, Chan K, Jarrard LE, Kanoski SE, Clegg DJ, Benoit SC. Contributions of the hippocampus and medial prefrontal cortex to energy and body weight regulation. *Hippocampus* **19**, 235-252 (2009).

26. Borcherding DC*, et al.* Dopamine receptors in human adipocytes: expression and functions. *PloS one* **6**, e25537 (2011).

27. Wiewiora M, Piecuch J, Sedek L, Mazur B, Sosada K. The effects of obesity on CD47 expression in erythrocytes. *Cytometry Part B, Clinical cytometry*, (2015).

28. Maimaitiyiming H, Norman H, Zhou Q, Wang S. CD47 deficiency protects mice from diet-induced obesity and improves whole body glucose tolerance and insulin sensitivity. *Scientific reports* **5**, 8846 (2015).

29. Forstermann U, Munzel T. Endothelial nitric oxide synthase in vascular disease: from marvel to menace. *Circulation* **113**, 1708-1714 (2006).

30. Holvoet P. Stress in obesity and associated metabolic and cardiovascular disorders. *Scientifica* **2012**, 205027 (2012).

31. Furukawa S*, et al.* Increased oxidative stress in obesity and its impact on metabolic syndrome. *The Journal of clinical investigation* **114**, 1752-1761 (2004).

32. Gasser JA, Inuzuka H, Lau AW, Wei W, Beroukhim R, Toker A. SGK3 mediates INPP4B-dependent PI3K signaling in breast cancer. *Molecular cell* **56**, 595-607 (2014).

33. Manning BD, Cantley LC. AKT/PKB signaling: navigating downstream. *Cell* **129**, 1261-1274 (2007).

34. Micu I*, et al.* NMDA receptors mediate calcium accumulation in myelin during chemical ischaemia. *Nature* **439**, 988-992 (2006).

35. Zhong HJ, Huo ZH, Dang J, Chen J, Zhu YS, Liu JH. Functional polymorphisms of the glutamate receptor N-methyl D-aspartate 2A gene are associated with heroin addiction. *Genetics and molecular research : GMR* **13**, 8714-8721 (2014).

36. Rezvani K, Teng Y, Shim D, De Biasi M. Nicotine regulates multiple synaptic proteins by inhibiting proteasomal activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 10508-10519 (2007).

37. Chen HS, Lipton SA. Pharmacological implications of two distinct mechanisms of interaction of memantine with N-methyl-D-aspartate-gated channels. *The Journal of pharmacology and experimental therapeutics* **314**, 961-971 (2005).

38. Maler JM*, et al.* Memantine inhibits ethanol-induced NMDA receptor up-regulation in rat hippocampal neurons. *Brain research* **1052**, 156-162 (2005).

39. Rogawski MA, Wenk GL. The neuropharmacological basis for the use of memantine in the treatment of Alzheimer's disease. *CNS drug reviews* **9**, 275-308 (2003).

40. Xu K, Lipsky RH. Repeated ketamine administration alters N-methyl-D-aspartic acid receptor subunit gene expression: implication of genetic vulnerability for ketamine abuse and ketamine psychosis in humans. *Exp Biol Med (Maywood)* **240**, 145-155 (2015).

41. Linden R, Martins VR, Prado MA, Cammarota M, Izquierdo I, Brentani RR. Physiology of the prion protein. *Physiological reviews* **88**, 673-728 (2008).

42. Redecke L*, et al.* Structural characterization of beta-sheeted oligomers formed on the pathway of oxidative prion protein aggregation in vitro. *Journal of structural biology* **157**, 308-320 (2007).

43. United States. Public Health Service. Office of the Surgeon General. *How tobacco smoke causes disease : the biology and behavioral basis for smoking-attributable disease : a report of the Surgeon General*. U.S. Dept. of Health and Human Services, Public Health Service (2010).

44. Bernhard D, Rossmann A, Wick G. Metals in cigarette smoke. *IUBMB life* **57**, 805-809 (2005).

45. Savini I, Rossi A, Pierro C, Avigliano L, Catani MV. SVCT1 and SVCT2: key proteins for vitamin C uptake. *Amino acids* **34**, 347-355 (2008).

46. Babaev VR, Li L, Shah S, Fazio S, Linton MF, May JM. Combined vitamin C and vitamin E deficiency worsens early atherosclerosis in apolipoprotein E-deficient mice. *Arteriosclerosis, thrombosis, and vascular biology* **30**, 1751-1757 (2010).

47. Bornstein SR*, et al.* Impaired adrenal catecholamine system function in mice with deficiency of the ascorbic acid transporter (SVCT2). *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **17**, 1928-1930 (2003).

48. Uhl GR, Drgon T, Li CY, Johnson C, Liu QR. Smoking and smoking cessation in disadvantaged women: assessing genetic contributions. *Drug and alcohol dependence* **104 Suppl 1**, S58-63 (2009).

49. Rose JE, Behm FM, Drgon T, Johnson C, Uhl GR. Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score. *Mol Med* **16**, 247-253 (2010).

50. Sasaki Y, Araki T, Milbrandt J. Stimulation of nicotinamide adenine dinucleotide biosynthetic pathways delays axonal degeneration after axotomy. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26**, 8484-8491 (2006).

51. Jayaram HN, Kusumanchi P, Yalowitz JA. NMNAT expression and its relation to NAD metabolism. *Current medicinal chemistry* **18**, 1962-1972 (2011).

52. Fang C, Decker H, Banker G. Axonal transport plays a crucial role in mediating the axon-protective effects of NmNAT. *Neurobiology of disease* **68**, 78-90 (2014).

53. Press C, Milbrandt J. Nmnat delays axonal degeneration caused by mitochondrial and oxidative stress. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28**, 4861-4871 (2008).

54. Warnatz HJ*, et al.* The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. *The Journal of biological chemistry* **286**, 23521-23532 (2011).

55. Liu JZ*, et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature genetics* **42**, 436-440 (2010).

56. Burdett T*, et al.* The NHGRI-EBI Catalog of published genome-wide association studies. (ed^(eds). v1.0 edn (2015).

57. Hindorff LA*, et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9362-9367 (2009).

58. Wood AR*, et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics* **46**, 1173-1186 (2014).

59. Eliasson B, Smith U. Leptin levels in smokers and long-term users of nicotine gum. *European journal of clinical investigation* **29**, 145-152 (1999).

60. Pisinger C, Jorgensen T. Waist circumference and weight following smoking cessation in a general population: the Inter99 study. *Preventive medicine* **44**, 290-295 (2007).

61. Day FR, Loh PR, Scott RA, Ong KK, Perry JR. A Robust Example of Collider Bias in a Genetic Association Study. *American journal of human genetics* **98**, 392-393 (2016).

62. Aschard H, Vilhjalmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *American journal of human genetics* **96**, 329-339 (2015).

63. Aschard H*, et al.* Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Human genetics* **131**, 1591-1613 (2012).

64. Winham SJ, Biernacka JM. Gene-environment interactions in genome-wide association studies: current approaches and new directions. *Journal of child psychology and psychiatry, and allied disciplines* **54**, 1120-1134 (2013).

65. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* **19**, 640-648 (2008).

66. Wei Y*, et al.* Chronic exposure to air pollution particles increases the risk of obesity and metabolic syndrome: findings from a natural experiment in Beijing. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, (2016).

67. Arany I, Hall S, Reed DK, Reed CT, Dixit M. Nicotine Enhances High-Fat Diet-Induced Oxidative Stress in the Kidney. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*, (2016).

68. Voight BF*, et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS genetics* **8**, e1002793 (2012).

69. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annual review of genomics and human genetics* **10**, 387-406 (2009).

70. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics* **5**, e1000529 (2009).

71. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics* **3**, e114 (2007).

72. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics* **81**, 1084-1097 (2007).

73. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology* **34**, 816-834 (2010).

74. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics* **39**, 906-913 (2007).

75. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC bioinformatics* **11**, 134 (2010).

76. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-1296 (2007).

77. Abecasis GR, Wigginton JE. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. *American journal of human genetics* **77**, 754-767 (2005).

78. Purcell S*, et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575 (2007).

79. Kutalik Z*, et al.* Methods for testing association between uncertain genotypes and quantitative traits. *Biostatistics* **12**, 1-17 (2011).

80. Winkler TW*, et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature protocols* **9**, 1192-1212 (2014).

81. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).

82. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

83. Randall JC*, et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS genetics* **9**, e1003500 (2013).

84. Winkler TW*, 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 genetics* **11**, e1005378 (2015).

85. Winkler TW, Kutalik Z, Gorski M, Lottaz C, Kronenberg F, Heid IM. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259-261 (2015).

86. Pruim RJ*, et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).

87. Abecasis GR*, et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073 (2010).

88. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research* **40**, D930-934 (2012).

89. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Briefings in bioinformatics* **14**, 144-161 (2013).

90. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580-585 (2013).

91. Boyle AP*, et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome research* **22**, 1790-1797 (2012).

92. Zhang X*, et al.* Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. *BMC genomics* **15**, 532 (2014).

93. Irizarry RA*, et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264 (2003).

94. Kutalik Z, Whittaker J, Waterworth D, Beckmann JS, Bergmann S. Novel method to estimate the phenotypic variation explained by genome-wide association studies reveals large fraction of the missing heritability. *Genetic epidemiology* **35**, 341-349 (2011).

95. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods* **10**, 5-6 (2013).

96. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**, 955-959 (2012).

ACKNOWLEDGEMENTS

A full list of acknowledgments appears in the Supplementary Note 4. Funding for this study was provided by the Aase and Ejner Danielsens Foundation; Academy of Finland (41071, 77299, 102318, 110413, 117787, 121584, 123885, 124243, 124282, 126925, 129378, 134309, 286284); Accare Center for Child and Adolescent Psychiatry; Action on Hearing Loss (G51); Agence Nationale de la 359 Recherche; Agency for Health Care Policy Research (HS06516); ALF/LUA research grant in Gothenburg; ALFEDIAM; ALK-Abello´ A/S; Althingi; American Heart Association (13POST16500011); Amgen; Andrea and Charles Bronfman Philanthropies; Ardix Medical; Arthritis Research UK; Association Diabète Risque Vasculaire; Australian National Health and Medical Research Council (241944, 339462, 389875, 389891, 389892, 389927, 389938, 442915, 442981, 496739, 552485, 552498); Avera Institute; Bayer Diagnostics; Becton Dickinson; BHF (RG/14/5/30893); Boston Obesity Nutrition Research Center (DK46200), Bristol-Myers Squibb; British Heart Foundation (RG/10/12/28456, RG2008/08, RG2008/014, SP/04/002); Medical Research Council of Canada; Canadian Institutes for Health Research (FRCN-CCT-83028); Cancer Research UK; Cardionics; Cavadis B.V., Center for Medical Systems Biology; Center of Excellence in Genomics; CFI; CIHR; City of Kuopio; CNAMTS; Cohortes Santé TGIR; Contrat de Projets État-Région; Croatian Science Foundation (8875); Danish Agency for Science, Technology and Innovation; Danish Council for Independent Research (DFF-1333-00124, DFF-1331-00730B); County Council of Dalarna; Dalarna University; Danish Council for Strategic Research; Danish Diabetes Academy; Danish Medical Research Council; Department of Health, UK; Development Fund from the University of Tartu (SP1GVARENG); Diabetes Hilfs- und Forschungsfonds Deutschland; Diabetes UK; Diabetes Research and Wellness Foundation Fellowship; Donald W. Reynolds Foundation; Dr. Robert Pfleger-Stiftung; Dutch Brain Foundation; Dutch Diabetes Research Foundation; Dutch Inter University Cardiology Institute; Dutch Kidney Foundation (E033); Dutch Ministry of Justice; the DynaHEALTH action No 633595, Economic Structure Enhancing Fund of the Dutch Government; Else Kröner-Fresenius-Stiftung

(2012_A147, P48/08//A11/08); Emil Aaltonen Foundation; Erasmus University Medical Center Rotterdam; Erasmus MC and Erasmus University Rotterdam; the Municipality of Rotterdam; Estonian Government (IUT20-60, IUT24-6); Estonian Research Roadmap through the Estonian Ministry of Education and Research (3.2.0304.11-0312); European Research Council (ERC Starting Grant and 323195:SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC); European Regional Development Fund; European Science Foundation (EU/QLRT-2001-01254); European Commission (018947, 018996, 201668, 223004, 230374, 279143, 284167, 305739, BBMRI-LPC-313010, HEALTH-2011.2.4.2-2-EU-MASCARA, HEALTH-2011-278913, HEALTH-2011-294713-EPLORE, HEALTH-F2-2008-201865-GEFOS, HEALTH-F2-2013-601456, HEALTH-F4-2007-201413, HEALTH-F4-2007-201550-HYPERGENES, HEALTH-F7-305507 HOMAGE, IMI/115006, LSHG-CT-2006-018947, LSHG-CT-2006-01947, LSHM-CT-2004-005272, LSHM-CT-2006-037697, LSHM-CT-2007-037273, QLG1-CT-2002-00896, QLG2-CT-2002-01254); Faculty of Biology and Medicine of Lausanne; Federal Ministry of Education and Research (01ZZ0103, 01ZZ0403, 01ZZ9603, 03IS2061A, 03ZIK012); Federal State of Mecklenburg-West Pomerania; Fédération Française de Cardiologie; Finnish Cultural Foundation; Finnish Diabetes Association; Finnish Foundation of Cardiovascular Research; Finnish Heart Association; Fondation Leducq; Food Standards Agency; Foundation for Strategic Research; French Ministry of Research; FRSQ; Genetic Association Information Network (GAIN) of the Foundation for the NIH; German Federal Ministry of Education and Research (BMBF, 01ER1206, 01ER1507); GlaxoSmithKline; Greek General Secretary of Research and Technology; Göteborg Medical Society; Health and Safety Executive; Healthcare NHS Trust; Healthway; Western Australia; Heart Foundation of Northern Sweden; Helmholtz Zentrum München – German Research Center for Environmental Health; Hjartavernd; Ingrid Thurings Foundation; INSERM; InterOmics (PB05 MIUR-CNR); INTERREG IV Oberrhein Program (A28); Interuniversity Cardiology Institute of the Netherlands (ICIN, 09.001); Italian Ministry of Health (ICS110.1/RF97.71); Italian Ministry of Economy and Finance (FaReBio di Qualità); Marianne and Marcus Wallenberg Foundation; the Ministry of Health,

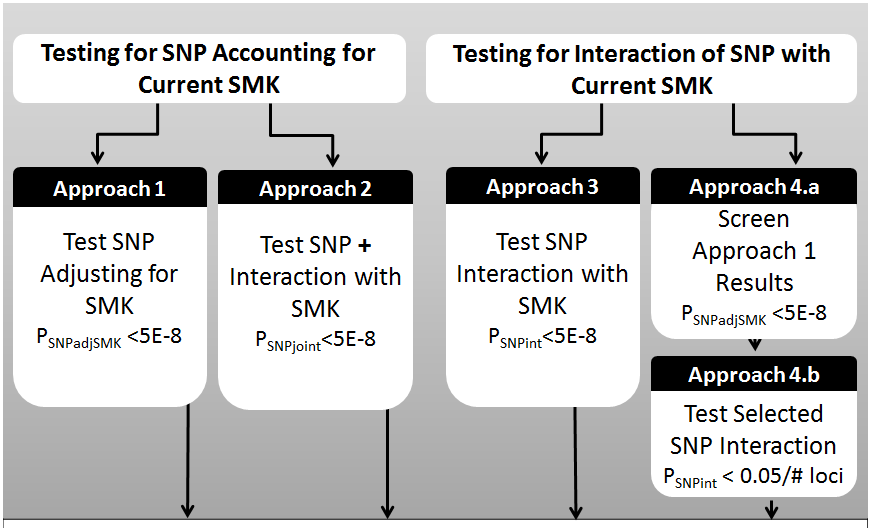
Welfare and Sports, the Netherlands; John D and Catherine T MacArthur Foundation Research Networks on Successful Midlife Development and Socioeconomic Status and Health; Juho Vainio Foundation; Juvenile Diabetes Research Foundation International; KfH Stiftung Präventivmedizin e.V.; King's College London; Knut and Alice Wallenberg Foundation; Kuopio University Hospital; Kuopio, Tampere and Turku University Hospital Medical Funds (X51001); La Fondation de France; Leenaards Foundation; Lilly; LMUinnovativ; Lundberg Foundation; Magnus Bergvall Foundation; MDEIE; Medical Research Council UK (G0000934, G0601966, G0700931, MC_U106179471, MC_UU_12019/1); MEKOS Laboratories; Merck Santé; Ministry for Health, Welfare and Sports, The Netherlands; Ministry of Cultural Affairs of Mecklenburg-West Pomerania; Ministry of Economic Affairs, The Netherlands; Ministry of Education and Culture of Finland (627;2004-2011); Ministry of Education, Culture and Science, The Netherlands; Ministry of Science, Education and Sport in the Republic of Croatia (108-1080315-0302); MRC centre for Causal Analyses in Translational Epidemiology; MRC Human Genetics Unit; MRC-GlaxoSmithKline pilot programme (G0701863); MSD Stipend Diabetes; National Institute for Health Research; Netherlands Brain Foundation (F2013(1)-28); Netherlands CardioVascular Research Initiative (CVON2011-19); Netherlands Genomics Initiative (050-060-810); Netherlands Heart Foundation (2001 D 032, NHS2010B280); Netherlands Organization for Scientific Research (NWO) and Netherlands Organisation for Health Research and Development (ZonMW) (56-464-14192, 60-60600-97-118, 100-001-004, 261-98-710, 400-05-717, 480-04-004, 480-05-003, 481-08-013, 904-61-090, 904-61-193, 911-11-025, 985-10-002, Addiction-31160008, BBMRI–NL 184.021.007, GB-MaGW 452-04-314, GB-MaGW 452-06-004, GB-MaGW 480-01-006, GB-MaGW 480-07-001, GB-MW 940-38-011, Middelgroot-911-09-032, NBIC/BioAssist/RK 2008.024, Spinozapremie 175.010.2003.005, 175.010.2007.006); Neuroscience Campus Amsterdam; NHS Foundation Trust; National Institutes of Health (1RC2MH089951, 1Z01HG000024, 24152, 263MD9164, 263MD821336, 2R01LM010098, 32100-2, 32122, 32108, AA07535, AA10248, AA11998, AA13320, AA13321, AA13326, AA14041, AA17688, AG13196, CA047988, DA12854,

DK56350, DK063491, DK078150, DK091718, DK100383, DK078616, ES10126, HG004790, HHSN268200625226C, HHSN268200800007C, HHSN268201200036C, HHSN268201500001I, HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, HHSN271201100004C, HL043851, HL45670, HL080467, HL085144, HL087660, HL054457, HL119443, HL118305, HL071981, HL034594, HL126024, HL130114, MH66206, MH081802, N01AG12100, N01HC55015, N01HC55016, N01C55018, N01HC55019, N01HC55020, N01HC55021, N01HC55022, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, N01HC95159, N01HC95160, N01HC95161, N01HC95162, N01HC95163, N01HC95164, N01HC95165, N01HC95166, N01HC95167, N01HC95168, N01HC95169, N01HG65403, N01WH22110, N02HL6-4278, N01-HC-25195, P01CA33619, R01AG023629, R01CA63, R01D004215701A, R01DK075787, R01DK062370, R01DK072193, R01DK075787, R01DK089256, R01HL53353, R01HL59367, R01HL086694, R01HL087641, R01HL087652, R01HL103612, R01HL105756, R01HL117078, R01HL120393, R03 AG046389, R37CA54281, RC2AG036495, RC4AG039029, RPPG040710371, RR20649, TW008288, TW05596, U01AG009740, U01CA98758, U01CA136792, U01DK062418, U01HG004402, U01HG004802, U01HG007376, U01HL080295, UL1RR025005, UL1TR000040, UL1TR000124, UL1TR001079, 2T32HL007055-36, T32GM074905, HG002651, HL084729, N01-HC-25195, UM1CA182913); NIH, National Institute on Aging (Intramural funding, NO1-AG-1-2109); Northern Netherlands Collaboration of Provinces; Novartis Pharma; Novo Nordisk; Novo Nordisk Foundation; ONIVINS; Parnassia Bavo group; Pierre Fabre; Province of Groningen; Päivikki and Sakari Sohlberg Foundation; Påhlssons Foundation; Paavo Nurmi Foundation; Radboud Medical Center Nijmegen; Research Centre for Prevention and Health, the Capital Region of Denmark; the Research Institute for Diseases in the Elderly; Research into Ageing; Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center; Roche; Royal Society; Russian Foundation for Basic Research (NWO-RFBR 047.017.043); Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06); Sanofi-Aventis;

1647 Scottish Government Health Directorates, Chief Scientist Office (CZD/16/6); Siemens Healthcare; Social
1648 Insurance Institution of Finland (4/26/2010); Social Ministry of the Federal State of Mecklenburg-West
1649 Pomerania; Société Francophone du 358 Diabète; State of Bavaria; Stiftelsen för Gamla Tjänarinnor;
1650 Stockholm County Council (560183, 592229); Strategic Cardiovascular and Diabetes Programmes of
1651 Karolinska Institutet and Stockholm County Council; Stroke Association; Swedish Diabetes Association;
1652 Swedish Diabetes Foundation (2013-024); Swedish Foundation for Strategic Research; Swedish Heart-
1653 Lung Foundation (20120197); Swedish Research Council (0593, 8691, 2012-1397, 2012-1727, and 2012-
1654 2215); Swedish Society for Medical Research; Swiss Institute of Bioinformatics; Swiss National Science
1655 Foundation (3100AO-116323/1, 31003A-143914, 33CSCO-122661, 33CS30-139468, 33CS30-148401,
1656 51RTP0_151019); Tampere Tuberculosis Foundation; Technology Foundation STW (11679); The Fonds
1657 voor Wetenschappelijk Onderzoek Vlaanderen, Ministry of the Flemish Community (G.0880.13,
1658 G.0881.13); The Great Wine Estates of the Margaret River Region of Western Australia; Timber
1659 Merchant Vilhelm Bangs Foundation; Topcon; Tore Nilsson Foundation; Torsten and Ragnar Söderberg's
1660 Foundation; United States – Israel Binational Science Foundation (Grant 2011036), Umeå University;
1661 University Hospital of Regensburg; University of Groningen; University Medical Center Groningen;
1662 University of Michigan; University of Utrecht; Uppsala Multidisciplinary Center for Advanced
1663 Computational Science (UPPMAX) (b2011036); Velux Foundation; VU University's Institute for Health
1664 and Care Research; Västra Götaland Foundation; Wellcome Trust (068545, 076113, 079895, 084723,
1665 088869, WT064890, WT086596, WT098017, WT090532, WT098051, 098381); Wissenschaftsoffensive
1666 TMO; Yrjö Jahnsson Foundation; and Åke Wiberg Foundation. This work was performed under the
1667 auspices of the Genetic Investigation of Anthropometric Traits (GIANT) Consortium. We acknowledge
1668 the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium for
1669 encouraging CHARGE studies to participate in this effort and for the contributions of CHARGE members
1670 to the analyses conducted for this research. The views expressed in this manuscript are those of the

1671 authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute
1672 (NHLBI); the National Institutes of Health (NIH); or the U.S. Department of Health and Human Services.
1673
1674 **COMPETING FINANCIAL INTERESTS**
1675 Bruce Psaty serves on the DSMB for a clinical trial funded by the device manufacturer (Zoll LifeCor) and
1676 on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson.

Figure 1. **Summary of study design and results**. Approach 1 uses both SNP and SMK in the association model. Approaches 2 and 3 use the SMK-stratified meta-analyses. Approach 4 screens loci based on Approach 1, then uses SMK-stratified results to identify loci with significant interaction effects (**Methods**).



Total Number of Significant Loci/Number of Novel

	Effect of SNP Accounting for Current SMK		Interaction of SNP with Current SMK	
Trait	Approach 1	Approach 2	Approach 3	Approach 4
BMI	65 / 4	57 / 2	2 / 1	1 / 0
WCadjBMI	72 / 9	66 / 5	2 / 2	0
WHRadjBMI	44 / 5	24 / 2	0	5 / 0

Total Non-Overlapping Loci

BMI: 68 / 6 **WCadjBMI: 76 / 11** **WHRadjBMI: 45 / 6**

Figure 2. **Forest plot for novel and GxSMK loci stratified by smoking status**. Estimated effect estimates (β ± 95% CI) per risk allele for **a**) BMI, **b**) WCadjBMI, and **c**) WHRadjBMI for novel loci from Approaches 1 and 2 (SNPadjSMK and SNPjoint, respectively) and all loci from Approaches 3 and 4 (SNPint and SNPscreen) identified in the primary meta-analyses. Loci are ordered by greater magnitude of effect in smokers compared to nonsmokers and labeled with the nearest gene. For the locus near *TMEM38B,* rs9409082 was used for effect estimates in this plot. (¥ loci identified for Approach 4, *loci identified for Approach 3).

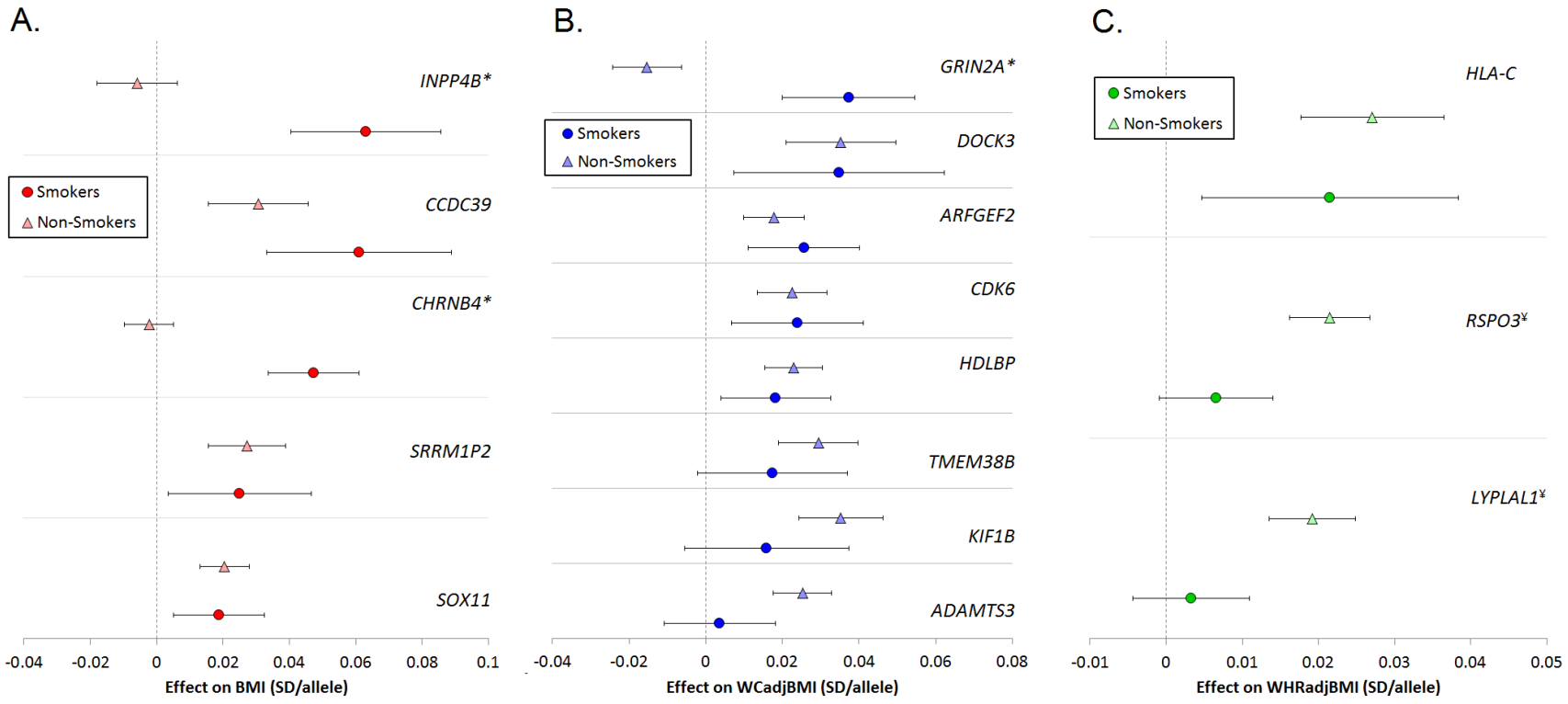
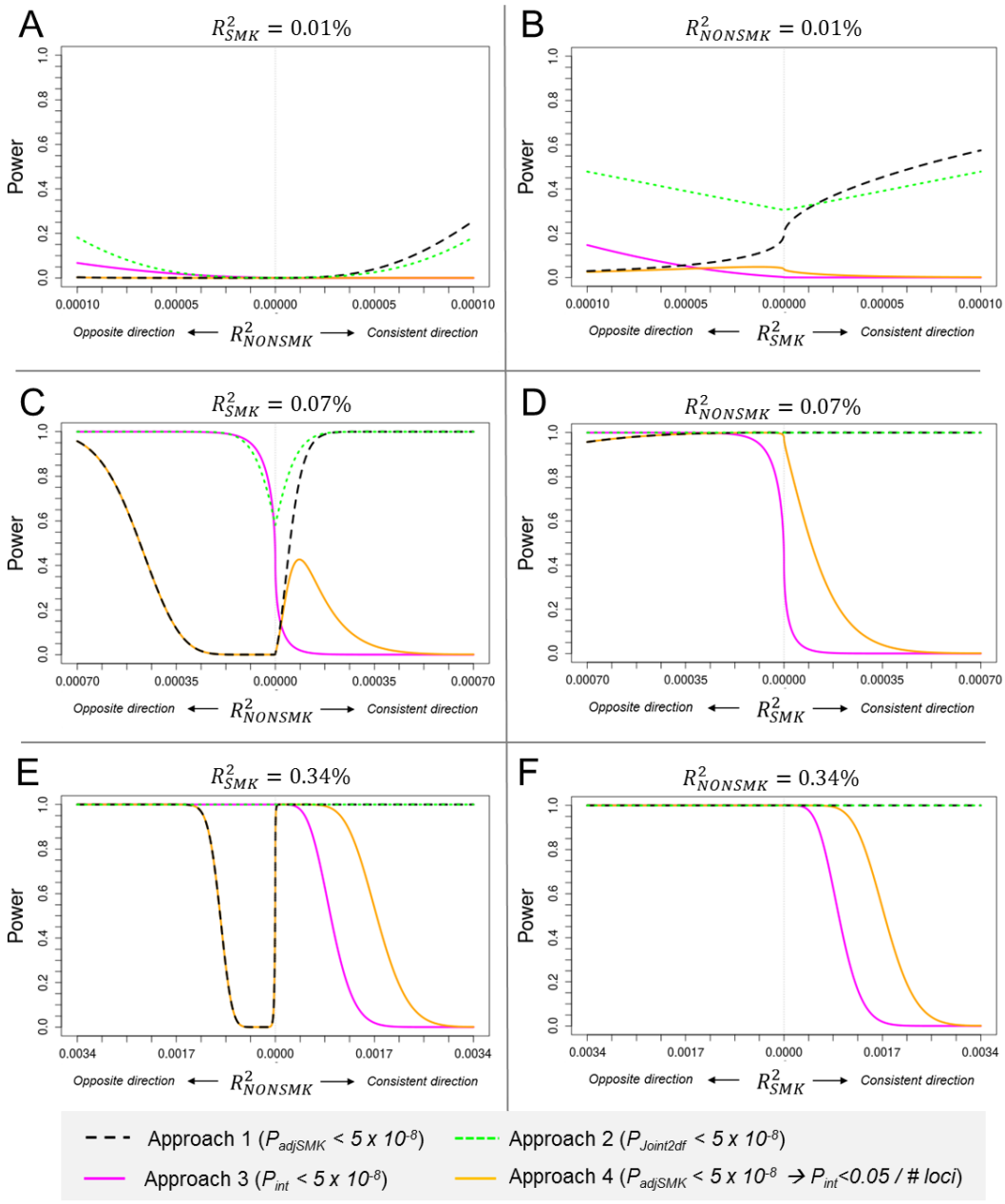
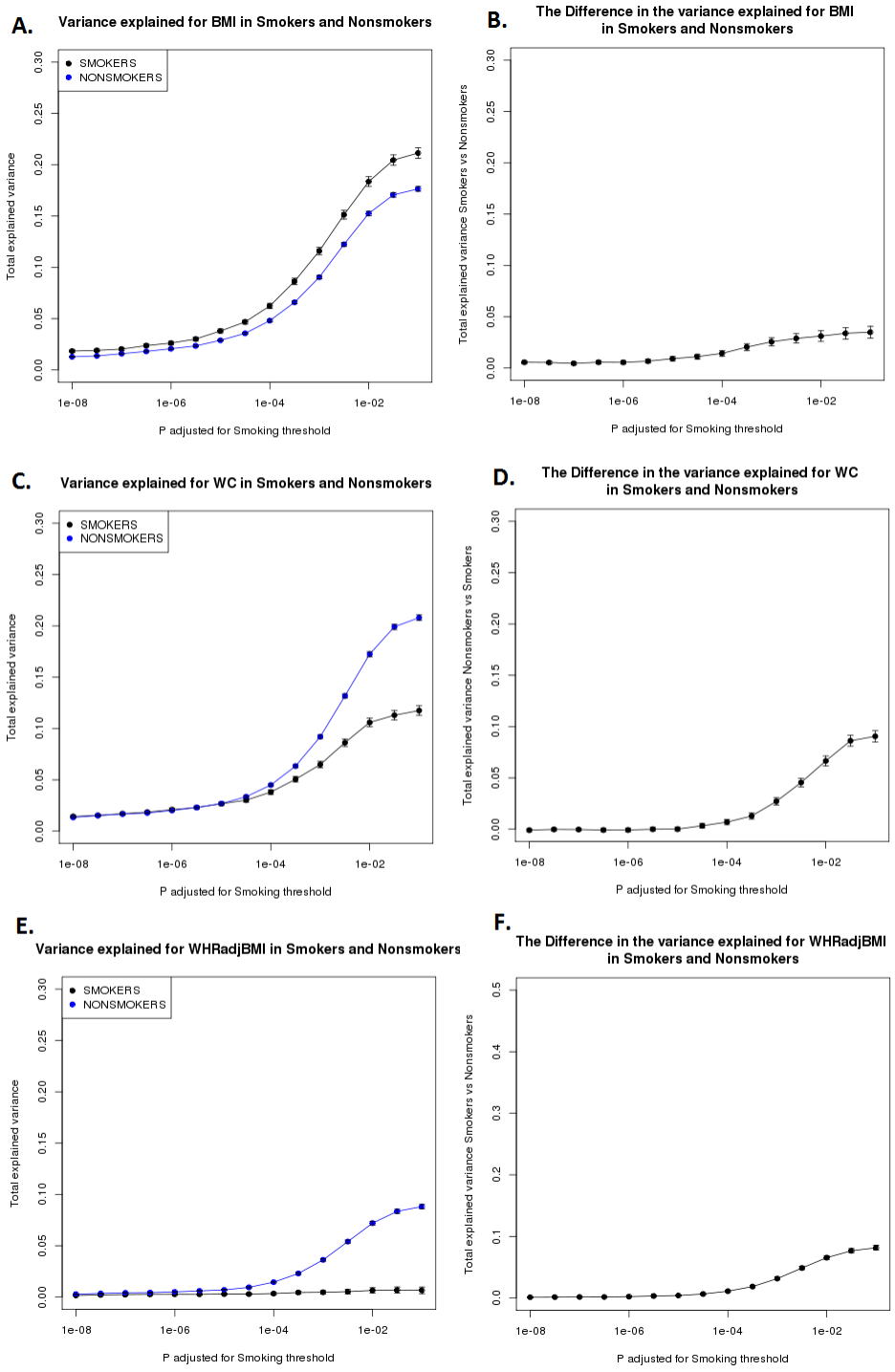


Figure 3. Power comparison across Approaches. Shown is the power to identify adjusted (Approach 1, dashed black lines), joint (Approach 2, dotted green lines) and interaction (Approach 3 and 4, solid magenta and orange lines) effects for various combinations of SMK- and NonSMK-specific effects and assuming 50,000 smokers and 180,000 nonsmokers. For Figures **a**, **c** and **e**, the effect in smokers was fixed at a small (R^2_{SMK}=0.01%, similar to the realistic *NUDT3* effect on BMI), medium (R^2_{SMK}=0.07%, similar to the realistic *BDNF* effect on BMI) or large (R^2_{SMK}=0.34%, similar to the realistic *FTO* effect on BMI) genetic effect, respectively, and varied in nonsmokers. For Figures **b**, **d** and **f**, the effect in nonsmokers was fixed to the small, medium and large BMI effects, respectively, and varied in smokers.


A
$R^2_{SMK} = 0.01\%$
B
$R^2_{NONSMK} = 0.01\%$
C
$R^2_{SMK} = 0.07\%$
D
$R^2_{NONSMK} = 0.07\%$
E
$R^2_{SMK} = 0.34\%$
F
$R^2_{NONSMK} = 0.34\%$
Power
Opposite direction
R^2_{NONSMK}
Consistent direction
R^2_{SMK}
Approach 1 ($P_{adjSMK} < 5 \times 10^{-8}$)
Approach 2 ($P_{Joint2df} < 5 \times 10^{-8}$)
Approach 3 ($P_{int} < 5 \times 10^{-8}$)
Approach 4 ($P_{adjSMK} < 5 \times 10^{-8} \rightarrow P_{int} < 0.05 / \#\ loci$)

1699
1700

Figure 4. **Stratum specific estimates of variance explained**. Total smoking status-specific explained variance (+/- SE) by SNPs meeting varying thresholds of overall association in Approach 1 (SNPadjSMK) and the difference between the proportion of variance explained between smokers and nonsmokers for these same sets of SNPs in BMI (**a,b**), WCadjBMI (**c,d**), and for WHRadjBMI (**e,f**).


A.
Variance explained for BMI in Smokers and Nonsmokers
SMOKERS
NONSMOKERS
Total explained variance
P adjusted for Smoking threshold
B.
The Difference in the variance explained for BMI in Smokers and Nonsmokers
Total explained variance Smokers vs Nonsmokers
P adjusted for Smoking threshold
C.
Variance explained for WC in Smokers and Nonsmokers
SMOKERS
NONSMOKERS
Total explained variance
P adjusted for Smoking threshold
D.
The Difference in the variance explained for WC in Smokers and Nonsmokers
Total explained variance Nonsmokers vs Smokers
P adjusted for Smoking threshold
E.
Variance explained for WHRadjBMI in Smokers and Nonsmokers
SMOKERS
NONSMOKERS
Total explained variance
P adjusted for Smoking threshold
F.
The Difference in the variance explained for WHRadjBMI in Smokers and Nonsmokers
Total explained variance Smokers vs Nonsmokers
P adjusted for Smoking threshold

1705
1706

Table 1. Summary of association results for novel loci reaching genome-wide significance in Approach (App) 1 ($P_{SNPadjSMK}$<5E-8) or Approach 2 ($P_{SNPjoint}$ <5E-8) for our primary meta-analysis in combined ancestries and combined sexes.

App	Marker	Chr:Pos (hg19)	Nearest Gene	N	EAF	Alleles E/O	SMOKERS β	SMOKERS P	NON-SMOKERS β	NON-SMOKERS P	Main and Interaction Effects β_{adj}	Main and Interaction Effects $P_{SNPadjSMK}$	Main and Interaction Effects P_{SNPint}	Main and Interaction Effects $P_{SNPjoint}$	GIANT + UKBB $P_{SNPadjSMK}$	GIANT + UKBB P_{SNPint}	GIANT + UKBB $P_{SNPjoint}$
BMI																	
1,2	rs10929925	2:6155557	*SOX11*	225,067	0.55	C/A	0.019	7.80E-03	0.02	8.40E-08	0.020	**1.1E-09**	8.2E-01	**1.6E-08**	**1.5E-13**	4.5E-01	**9.8E-13**
1	rs6794880	3:84451512	*SRRM1P2*	186,968	0.85	A/G	0.025	2.30E-02	0.027	3.90E-06	0.028	**4.3E-08**	8.5E-01	1.8E-06	**4.9E-09**	4.5E-01	9.7E-08
2	rs13069244	3:180441172	*CCDC39*	233,776	0.08	A/G	0.061	1.80E-05	0.031	6.60E-05	0.035	1.2E-07	4.6E-02	**3.5E-08**	**6.1E-10**	1.1E-02	**9.6E-11**
WCadjBMI																	
1,2	rs17396340	1:10286176	*KIF1B*	206,485	0.14	A/G	0.016	1.40E-01	0.035	**4.70E-10**	0.028	**3.0E-08**	9.8E-02	**9.1E-10**	**1.0E-11**	2.9E-02	**1.5E-13**
1,2	rs6743226	2:242236972	*HDLBP*	200,666	0.53	C/T	0.018	1.30E-02	0.023	**2.60E-09**	0.022	**1.2E-10**	5.5E-01	**5.8E-10**	**6.7E-12**	7.0E-01	**2.8E-11**
1	rs4378999	3:51208646	*DOCK3*	156,566	0.13	T/A	0.035	1.30E-02	0.035	1.30E-06	0.036	**4.1E-08**	9.7E-01	4.1E-07	**7.6E-11**	5.3E-01	**3.2E-10**
1,2	rs7697556	4:73515313	*ADAMTS3*	206,017	0.49	T/C	0.004	6.30E-01	0.025	**7.30E-11**	0.021	**5.2E-09**	6.7E-03	**7.6E-10**	**5.4E-19**	1.9E-02	**2.7E-19**
1	rs10269774	7:92253972	*CDK6*	157,552	0.34	A/G	0.024	6.60E-03	0.023	1.10E-06	0.023	**2.9E-08**	8.8E-01	1.6E-07	**2.9E-10**	7.7E-01	**2.1E-09**
1	rs6470765	8:130736697	*GSDMC*	157,450	0.76	A/C	0.032	1.90E-03	0.023	1.70E-05	0.026	**4.8E-08**	4.3E-01	9.5E-07	**2.5E-12**	8.9E-01	**9.0E-11**
2	rs9408815	9:108890521	*TMEM38B*	156,427	0.75	C/G	0.012	2.30E-01	0.03	**4.20E-09**	0.026	**2.3E-08**	8.5E-02	**1.7E-08**	**1.2E-11**	3.0E-01	**2.8E-11**
1	rs9409082	9:108901049		157,785	0.76	C/T	0.017	8.10E-02	0.029	**2.60E-08**	0.027	**1.5E-08**	2.7E-01	**4.6E-08**	**9.5E-12**	6.6E-01	**6.5E-11**
1	rs6012558	20:47531286	*ARFGEF2*	208,004	0.41	A/G	0.026	5.40E-04	0.018	6.50E-06	0.020	**1.9E-08**	3.3E-01	1.3E-07	**1.5E-09**	7.0E-02	**3.0E-09**
WHRadjBMI																	
1,2	rs1049281	6:31236567	*HLA-C*	149,285	0.66	C/T	0.022	1.30E-02	0.027	**2.00E-08**	0.025	**2.2E-09**	5.6E-01	**5.3E-09**	**1.2E-18**	8.3E-01	**1.8E-10**

Abbreviations: Chr- chromosome; Pos- position (bp); E/O- effect/other; EAF- effect allele frequency; adj- adjusted for smoking; int- interaction; App- Approach.

Table 2. Novel loci showing significant association in Approaches 1 (SNPadjSMK), 2 (SNPjoint), 3 (SNPint), and 4 (SNPscreen) for loci identified in secondary analysis samples, which were not identified in primary meta-analyses. All estimates are from the stratum specified in the Approach:Sample column (E-European-only, A- all ancestries, C- combined sexes, W-women only, M- men only). * This locus was filtered from approaches 2-4 due to low sample size in the SMK strata, and only p-values for Approach 1 are considered significant.

Approach: Strata	Marker	Chr:Pos (hg19)	Nearest Gene	N	EAF	Alleles E/O	SMOKERS β	SMOKERS P	NON-SMOKERS β	NON-SMOKERS P	Main and Interaction Effects β_{adj}	P_{adj}	P_{int}	P_{joint}	GIANT + UKBB $P_{SNPadjSMK}$	P_{SNPint}	$P_{SNPjoint}$
BMI																	
1:EC	rs2481665	1:62594677	*INADL*	209,453	0.56	T/C	0.015	4.60E-02	0.021	8.90E-08	0.019	**3.50E-08**	4.00E-01	6.70E-08	**3.3E-11**	7.8E-01	**2.0E-08**
1:AW	rs12629427	3:89145340	*EPHA3*	137,961	0.26	C/T	0.025	2.10E-02	0.028	3.60E-07	0.027	**4.80E-08**	8.00E-01	2.00E-07	7.7E-08	9.1E-01	3.0E-07
1:EW	rs2173039	3:89142175		117,942	0.26	C/G	0.024	3.10E-02	0.032	8.90E-08	0.031	**7.30E-09**	5.70E-01	6.50E-08	**2.4E-09**	9.3E-01	2.2E-07
WCadjBMI																	
1:EM	rs1545348	5:34718343	*RAI14*	77,677	0.73	T/G	0.044	3.10E-04	0.03	1.90E-05	0.034	**1.80E-08**	3.20E-01	1.70E-07	1.2E-07	1.2E-01	4.8E-07
2:EW	rs6076699	20:4566688	*PRNP*	76,930	0.97	A/G	0.169	1.40E-05	-0.07	1.20E-04	-0.034	3.50E-02	**1.40E-08**	**4.80E-08**	4.2E-02	2.3E-06	3.4E-06
WHRadjBMI																	
1:AW	rs670752	3:107312980	*BBX*	107,568	0.32	A/G	0.012	5.50E-02	0.009	1.50E-02	0.027	**4.90E-08**	6.80E-01	7.80E-03	**3.1E-10**	3.8E-01	9.5E-05
1:EC	rs589428	6:31848220	*EHMT2*	162,918	0.66	G/T	0.006	1.20E-01	0.011	4.10E-04	0.022	**2.80E-08**	3.50E-01	7.00E-04	**1.1E-17**	8.4E-02	**1.6E-10**
2:EC	rs1856293	6:133480940	*EYA4*	127,431	0.52	A/C	0.006	5.30E-01	-0.028	9.10E-09	-0.019	6.50E-06	5.40E-04	**4.70E-08**	9.6E-08	1.3E-02	**1.5E-08**
1:AW	rs2001945	8:126477978	*TRIB1*	103,446	0.4	G/C	0.009	1.20E-01	0.013	1.00E-04	0.025	**4.70E-08**	5.90E-01	1.30E-04	**1.1E-09**	3.0E-01	1.4E-06
1:EC	rs17065323	13:44627788	*SMIM2**	69,968	0.01	T/C	0.154	1.90E-01	-0.23	1.20E-10	-0.181	**9.20E-09**	1.40E-03	3.90E-10	**9.6E-09**	3.6E-03	**1.3E-09**

Abbreviations: Chr- chromosome, Pos- position (bp), E/O- effect/other, EAF- effect allele frequency, Padj- adjusted for smoking, int- interaction.

Table 3. Summary of association results for loci showing significance for interaction with smoking in Approach (App) 3 (SNPint) and/or Approach 4 (SNPscreen) in our primary meta-analyses of combined ancestries and combined sexes. Ł - known locus.

App	Marker	Chr:Pos (hg19)	Nearest Gene	N	EAF	Alleles E/O	SMOKERS β	SMOKERS P	NON-SMOKERS β	NON-SMOKERS P	Main and Interaction Effects β_{adj}	P_{adj}	P_{int}	P_{joint}	GIANT + UKBB $P_{SNPadjSMK}$	P_{SNPint}	$P_{SNPjoint}$
BMI																	
3	rs336396	4:143062811	*INPP4B*	169,646	0.18	T/C	0.063	**4.8E-08**	-0.006	3.4E-01	0.007	2.3E-01	**2.1E-08**	1.9E-07	7.4E-01	2.7E-06	1.3E-05
3	rs12902602 Ł	15:78967401	*CHRNB4*	240,135	0.62	A/G	0.047	**1.8E-11**	-0.002	5.5E-01	0.009	8.6E-03	**4.1E-11**	**1.1E-10**	1.1E-01	**6.0E-13**	**1.6E-12**
WCadjBMI																	
3	rs4141488	16:9629067	*GRIN2A*	153,892	0.5	T/C	0.037	2.2E-05	-0.015	9.6E-04	-0.003	4.4E-01	**2.7E-08**	5.0E-07	9.5E-01	1.8E-06	1.1E-05
WHRadjBMI																	
4	rs765751 Ł	1:219669226	*LYPLAL1*	189,028	0.64	C/T	0.003	3.9E-01	0.019	**3.1E-11**	0.029	**3.1E-16**	*7.3E-04*	**2.1E-10**	**9.1E-31**	*1.4E-04*	**7.8E-22**
4	rs7766106 Ł	6:127455138	*RSPO3*	188,174	0.48	T/C	0.007	7.9E-02	0.022	**2.2E-15**	0.037	**3.7E-27**	*9.7E-04*	**3.8E-15**	**4.4E-51**	*1.0E-05*	**3.4E-34**

Abbreviations: Chr- chromosome; Pos- position (bp); E/O- effect/other; EAF- effect allele frequency; adj- adjusted for smoking; int- interaction; App- Approach.

Table 4. Summary of association results for loci showing significance for interaction with smoking in Approach 3 (SNPint) and/or Approach 4 (SNPscreen) in our secondary meta-analyses not identified in primary meta-analyses. All estimates are from the stratum specified in the Approach:Sample column (E-European-only, A- all ancestries, C- combined sexes, W-women only, M- men only). Ł - known locus. The R^2 between the *ADAMTS7* (rs1809420) and *CHRNB4* variant (rs1290362) in **Table 3** is 0.72 (HapMap 2, CEU). Additionally, the *PRNP* variant (rs6076699) is the same as the variant that came up from Approach 2 (**Table 2**).

Approach: Strata	Marker	Chr:Pos (hg19)	Nearest Gene	N	EAF	Alleles E/O	SMOKERS β	SMOKERS P	NON-SMOKERS β	NON-SMOKERS P	Main and Interaction Effects β_{adj}	P_{adj}	P_{int}	P_{joint}	GIANT + UKBB $P_{SNPadjSMK}$	P_{SNPint}	$P_{SNPjoint}$
BMI																	
4:AM	rs1809420 Ł	15:79056769	*ADAMTS7*	57,081	0.59	T/C	0.074	**9.8E-08**	0.023	2.0E-03	0.036	**4.9E-08**	*9.4E-04*	**5.6E-09**	9.8E-05	*3.3E-05*	1.9E-07
WCadjBMI																	
3:EW	rs6076699	20:4566688	*PRNP*	76,930	0.97	A/G	0.169	1.4E-05	-0.07	1.2E-04	-0.034	3.5E-02	**1.4E-08**	**4.8E-08**	4.2E-02	2.3E-06	3.4E-06
WHRadjBMI																	
4:EM	rs30000 Ł	5:55803533	*MAP3K1*	71,424	0.27	G/A	0.002	7.8E-01	0.031	**3.7E-08**	0.04	**1.7E-10**	*1.6E-04*	2.7E-07	**2.7E-17**	*3.2E-07*	**3.8E-15**
4:AM	rs459193 Ł	5:55806751		80,852	0.27	A/G	0.004	5.0E-01	0.034	**4.1E-10**	0.043	**2.3E-13**	*6.8E-05*	**2.2E-09**	**3.5E-20**	*2.5E-07*	**1.6E-17**
4:AM	rs2071449 Ł	12:54428011	*HOXC4-*	70,868	0.37	A/C	0.003	6.0E-01	0.026	1.0E-06	0.034	**9.1E-09**	*1.1E-03*	5.7E-06	**2.7E-12**	8.0E-04	**2.8E-09**
4:EM	rs754133 Ł	12:54418920	*HOXC6*	71,136	0.36	A/G	0.003	6.2E-01	0.026	8.2E-07	0.034	**3.0E-09**	*1.1E-03*	4.0E-06	**2.1E-12**	*9.7E-04*	**4.0E-09**
4:AM	rs12608504 Ł	19:18389135	*JUND*	80,087	0.37	A/G	0.006	2.6E-01	0.025	5.0E-07	0.032	**4.7E-09**	*5.5E-03*	1.8E-06	**2.9E-11**	1.3E-02	**1.6E-08**

Abbreviations: E/O- effect/other, EAF- effect allele frequency, SE- standard error; Chr- chromosome; Pos- position (bp); adj- adjusted for smoking; int- interaction; App- Approach.