

Uppsala University

This is an accepted version of a paper published in *Nature Genetics*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:

Berndt, S., Gustafsson, S., Maegi, R., Ganna, A., Wheeler, E. et al. (2013)

"Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture"

Nature Genetics, 45(5): 501-U69

Access to the published version may require subscription.

DOI: 10.1038/ng.2606

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-200675>

DiVA 

<http://uu.diva-portal.org>

Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture

Sonja I. Berndt^{1*}, Stefan Gustafsson^{2,3*}, Reedik Mägi^{4,5*}, Andrea Ganna^{3*}, Eleanor Wheeler⁶, Mary F. Feitosa⁷, Anne E. Justice⁸, Keri L. Monda^{8,9}, Damien C. Croteau-Chonka¹⁰, Felix R. Day¹¹, Tõnu Esko^{5,12}, Tove Fall³, Teresa Ferreira⁴, Davide Gentilini¹³, Anne U. Jackson¹⁴, Jian'an Luan¹¹, Joshua C. Randall^{4,6}, Sailaja Vedantam^{15,16,17}, Cristen J. Willer^{18,19,20}, Thomas W. Winkler²¹, Andrew R. Wood²², Tsegaselassie Workalemahu^{23,24}, Yi-Juan Hu²⁵, Sang Hong Lee²⁶, Liming Liang^{27,28}, Dan-Yu Lin²⁹, Josine L. Min⁴, Benjamin M. Neale³⁰, Gudmar Thorleifsson³¹, Jian Yang^{32,33}, Eva Albrecht³⁴, Najaf Amin³⁵, Jennifer L. Bragg-Gresham¹⁴, Gemma Cadby^{36,37,38}, Martin den Heijer³⁹, Niina Eklund^{40,40}, Krista Fischer⁵, Anuj Goel⁴¹, Jouke-Jan Hottenga⁴², Jennifer E. Huffman⁴³, Ivonne Jarick⁴⁴, Åsa Johansson^{45,46}, Toby Johnson^{47,48}, Stavroula Kanoni⁶, Marcus E. Kleber^{49,50}, Inke R. König⁵¹, Kati Kristiansson⁴⁰, Zoltán Kutalik^{52,53}, Claudia Lamina⁵⁴, Cecile Lecoeur^{55,56}, Guo Li⁵⁷, Massimo Mangino⁵⁸, Wendy L. McArdle⁵⁹, Carolina Medina-Gomez^{35,60,61}, Martina Müller-Nurasyid^{34,62,63}, Julius S. Ngwa⁶⁴, Ilja M. Nolte⁶⁵, Lavinia Paternoster⁶⁶, Sonali Pechlivanis⁶⁷, Markus Perola^{5,40,68}, Marjolein J. Peters^{35,60,61}, Michael Preuss^{51,69}, Lynda M. Rose⁷⁰, Jianxin Shi¹, Dmitry Shungin^{71,72,73}, Albert Vernon Smith^{74,75}, Rona J. Strawbridge⁷⁶, Ida Surakka^{40,68}, Alexander Teumer⁷⁷, Mieke D. Trip^{78,79}, Jonathan Tyrer⁸⁰, Jana V. Van Vliet-Ostaptchouk^{81,82}, Liesbeth Vandenput⁸³, Lindsay L. Waite⁸⁴, Jing Hua Zhao¹¹, Devin Absher⁸⁴, Folkert W. Asselbergs⁸⁵, Mustafa Atalay⁸⁶, Antony P. Attwood⁸⁷, Anthony J. Balmforth⁸⁸, Hanneke Basart⁷⁸, John Beilby^{89,90}, Lori L. Bonnycastle⁹¹, Paolo Brambilla⁹², Marcel Bruinenberg⁸², Harry Campbell⁹³, Daniel I. Chasman^{70,94}, Peter S. Chines⁹¹, Francis S. Collins⁹¹, John M. Connell^{95,96}, William Cookson⁹⁷, Ulf de Faire⁹⁸, Femmie de Vegt⁹⁹, Mariano Dei¹⁰⁰, Maria Dimitriou¹⁰¹, Sarah Edkins⁶, Karol Estrada^{35,60,61}, David M. Evans⁶⁶, Martin Farrall⁴¹, Marco M. Ferrario¹⁰², Jean Ferrières¹⁰³, Lude Franke^{82,104}, Francesca Frau¹⁰⁵, Pablo V. Gejman^{106,107}, Harald Grallert¹⁰⁸, Henrik Grönberg³, Vilmondur Gudnason^{74,75}, Alistair S. Hall¹⁰⁹, Per Hall³, Anna-Liisa Hartikainen¹¹⁰, Caroline Hayward⁴³, Nancy L. Heard-Costa¹¹¹, Andrew C. Heath¹¹², Johannes Hebebrand¹¹³, Georg Homuth⁷⁷, Frank B. Hu²³, Sarah E. Hunt⁶, Elina Hyppönen¹¹⁴, Carlos Iribarren¹¹⁵, Kevin B. Jacobs^{1,116}, John-Olov Jansson¹¹⁷, Antti Jula¹¹⁸, Mika Kähönen¹¹⁹, Sekar Kathiresan^{120,121,122}, Frank Kee¹²³, Kay-Tee Khaw¹²⁴, Mika Kivimäki¹²⁵, Wolfgang Koenig¹²⁶, Aldi T.

Kraja⁷, Meena Kumari¹²⁵, Kari Kuulasmaa¹²⁷, Johanna Kuusisto¹²⁸, Jaana H. Laitinen¹²⁹, Timo A. Lakka^{86,130}, Claudia Langenberg^{11,125}, Lenore J. Launer¹³¹, Lars Lind¹³², Jaana Lindström¹³³, Jianjun Liu¹³⁴, Antonio Liuzzi¹³⁵, Marja-Liisa Lokki¹³⁶, Mattias Lorentzon⁸³, Pamela A. Madden¹¹², Patrik K. Magnusson³, Paolo Manunta¹³⁷, Diana Marek^{52,53}, Winfried März^{50,138}, Irene Mateo Leach¹³⁹, Barbara McKnight¹⁴⁰, Sarah E. Medland³³, Evelin Mihailov^{5,12}, Lili Milani⁵, Grant W. Montgomery³³, Vincent Mooser¹⁴¹, Thomas W. Mühleisen^{142,143}, Patricia B. Munroe^{47,48}, Arthur W. Musk^{144,145,146}, Narisu Narisu⁹¹, Gerjan Navis¹⁴⁷, George Nicholson^{148,149}, Ellen A. Nohr¹⁵⁰, Ken K. Ong^{11,151}, Ben A. Oostra^{61,152,153}, Colin N.A. Palmer¹⁵⁴, Aarno Palotie^{6,68}, John F. Peden¹⁵⁵, Nancy Pedersen³, Annette Peters^{108,156,157}, Ozren Polasek¹⁵⁸, Anneli Pouta^{110,159}, Peter P. Pramstaller^{160,161,162}, Inga Prokopenko^{4,163}, Carolin Pütter⁶⁷, Aparna Radhakrishnan^{6,164,165}, Olli Raitakari^{166,167}, Augusto Rendon^{87,164,165,168}, Fernando Rivadeneira^{35,60,61}, Igor Rudan⁹³, Timo E. Saaristo^{169,170}, Jennifer G. Sambrook^{164,165}, Alan R. Sanders^{106,107}, Serena Sanna¹⁰⁰, Jouko Saramies¹⁷¹, Sabine Schipf¹⁷², Stefan Schreiber¹⁷³, Heribert Schunkert^{69,174}, So-Youn Shin⁶, Stefano Signorini¹⁷⁵, Juha Sinisalo¹⁷⁶, Boris Skrobek^{55,56}, Nicole Soranzo^{6,58}, Alena Stančáková¹⁷⁷, Klaus Stark¹⁷⁸, Jonathan C. Stephens^{164,165}, Kathleen Stirrups⁶, Ronald P. Stolk^{65,82}, Michael Stumvoll^{179,180}, Amy J. Swift⁹¹, Eirini V. Theodoraki¹⁰¹, Barbara Thorand¹⁵⁶, David-Alexandre Tregouet¹⁸¹, Elena Tremoli¹⁸², Melanie M. Van der Klauw^{81,82}, Joyce B.J. van Meurs^{35,60,61}, Sita H. Vermeulen^{99,183}, Jorma Viikari¹⁸⁴, Jarmo Virtamo¹²⁷, Veronique Vitart⁴³, Gérard Waeber¹⁸⁵, Zhaoming Wang^{1,116}, Elisabeth Widén⁶⁸, Sarah H. Wild⁹³, Gonneke Willemsen⁴², Bernhard R. Winkelmann¹⁸⁶, Jacqueline C.M. Witteman^{35,61}, Bruce H.R. Wolffenbuttel^{81,82}, Andrew Wong¹⁵¹, Alan F. Wright⁴³, M. Carola Zillikens^{60,61}, Philippe Amouyel¹⁸⁷, Bernhard O. Boehm¹⁸⁸, Eric Boerwinkle¹⁸⁹, Dorret I. Boomsma⁴², Mark J. Caulfield^{47,48}, Stephen J. Chanock¹, L. Adrienne Cupples⁶⁴, Daniele Cusi^{105,190}, George V. Dedoussis¹⁰¹, Jeanette Erdmann^{69,174}, Johan G. Eriksson^{191,192,193}, Paul W. Franks^{71,72,194}, Philippe Froguel^{55,56,195}, Christian Gieger³⁴, Ulf Gyllensten⁴⁵, Anders Hamsten⁷⁶, Tamara B. Harris¹³¹, Christian Hengstenberg¹⁷⁸, Andrew A. Hicks¹⁶⁰, Aroon Hingorani¹²⁵, Anke Hinney¹¹³, Albert Hofman^{35,61}, Kees G. Hovingh⁷⁸, Kristian Hveem¹⁹⁶, Thomas Illig^{108,197}, Marjo-Riitta Jarvelin^{159,198,199,200}, Karl-Heinz Jöckel⁶⁷, Sirkka M. Keinanen-Kiukaanniemi^{199,201}, Lambertus A. Kiemeny^{99,202,203}, Diana Kuh¹⁵¹, Markku Laakso¹²⁸, Terho Lehtimäki²⁰⁴, Douglas F. Levinson²⁰⁵, Nicholas G. Martin³³, Andres Metspalu^{5,12}, Andrew

D. Morris¹⁵⁴, Markku S. Nieminen¹⁷⁶, Inger Njølstad^{206,207}, Claes Ohlsson⁸³, Albertine J. Oldehinkel²⁰⁸, Willem H. Ouwehand^{6,87,164,165}, Lyle J. Palmer^{36,37}, Brenda Penninx²⁰⁹, Chris Power¹¹⁴, Michael A. Province⁷, Bruce M. Psaty^{57,210,211}, Lu Qi^{23,24}, Rainer Rauramaa^{130,212}, Paul M. Ridker^{70,94}, Samuli Ripatti^{6,40,68}, Veikko Salomaa¹²⁷, Nilesh J. Samani^{213,214}, Harold Snieder^{65,82}, Thorkild I.A. Sørensen²¹⁵, Timothy D. Spector⁵⁸, Kari Stefansson^{31,216}, Anke Tönjes^{179,180}, Jaakko Tuomilehto^{133,217,218,219}, André G. Uitterlinden^{35,60,61}, Matti Uusitupa^{220,221}, Pim van der Harst^{104,139}, Peter Vollenweider¹⁸⁵, Henri Wallaschofski²²², Nicholas J. Wareham¹¹, Hugh Watkins⁴¹, H.-Erich Wichmann^{223,224,225}, James F. Wilson⁹³, Goncalo R. Abecasis¹⁴, Themistocles L. Assimes²²⁶, Inês Barroso^{6,227}, Michael Boehnke¹⁴, Ingrid B. Borecki⁷, Panos Deloukas⁶, Caroline S. Fox²²⁸, Timothy Frayling²², Leif C. Groop²²⁹, Talin Haritunian²³⁰, Iris M. Heid^{21,223}, David Hunter^{23,24}, Robert C. Kaplan²³¹, Fredrik Karpe^{163,232}, Miriam Moffatt⁹⁷, Karen L. Mohlke¹⁰, Jeffrey R. O'Connell²³³, Yudi Pawitan³, Eric E. Schadt^{234,235}, David Schlessinger²³⁶, Valgerdur Steinthorsdottir³¹, David P. Strachan²³⁷, Unnur Thorsteinsdottir^{31,216}, Cornelia M. van Duijn^{35,61,238}, Peter M. Visscher^{26,32}, Anna Maria Di Blasio¹³, Joel N. Hirschhorn^{15,16,17}, Cecilia M. Lindgren⁴, Andrew P. Morris⁴, David Meyre^{55,56,239}, André Scherag⁶⁷, Mark I. McCarthy^{4,163,232*}, Elizabeth K. Speliotes^{240,241*}, Kari E. North^{8*}, Ruth J.F. Loos^{11,242,243,244*}, Erik Ingelsson^{2,3,4*}

* These authors contributed equally to this work

Affiliations

A full list of author affiliations appears at the end of the paper.

Correspondence should be addressed to:

Erik Ingelsson, MD, PhD, FAHA
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
P.O Box 281, Solna, SE-171 77 Stockholm, SWEDEN
Phone: +46-8-52482334, fax: +46-8-314975
E-mail: erik.ingelsson@ki.se

Word count: 125 (abstract); 4,006 (main text).

Abstract

Approaches exploiting extremes of the trait distribution may reveal novel loci for common traits, but it is unknown whether such loci are generalizable to the general population. In a genome-wide search for loci associated with upper vs. lower 5th percentiles of body mass index, height and waist-hip ratio, as well as clinical classes of obesity including up to 263,407 European individuals, we identified four new loci (*IGFBP4*, *H6PD*, *RSRC1*, *PPP2R2A*) influencing height detected in the tails and seven new loci (*HNF4G*, *RPTOR*, *GNAT2*, *MRPS33P4*, *ADCY9*, *HS6ST3*, *ZZZ3*) for clinical classes of obesity. Further, we show that there is large overlap in terms of genetic structure and distribution of variants between traits based on extremes and the general population and little etiologic heterogeneity between obesity subgroups.

Twin studies have established a strong heritable component to body mass index (BMI; $h^2 \sim 40-70\%$),^{1,2} and height ($h^2 \sim 70-90\%$).³ Previous meta-analyses of genome-wide association studies (GWAS) have identified 36 genetic loci associated with BMI,⁴⁻⁶ 14 loci with waist-hip ratio adjusted for BMI (WHR) reflecting fat distribution,^{7,8} and 180 loci with height⁹, and contributed to our understanding of the genetic architecture of complex traits. However, established loci for complex traits only account for a small proportion of trait heritability, as discussed recently.^{10,11} Some postulated explanations for this include undiscovered low frequency variants with larger effects, imperfect tagging of causal variants, epistasis, gene-environment interaction, and phenotype heterogeneity. This has led to increasing interest in approaches exploiting extremes of the trait distribution, where there may be less locus heterogeneity, greater genetic contribution, and enrichment for highly penetrant variants. Utilization of extremes has also been proposed to improve cost-efficiency, since effect sizes may be larger, fewer subjects may be needed for genotyping, and a smaller proportion of the variance may be attributable to environmental factors. Indeed, several prior studies have used extreme designs to discover novel loci for various complex traits, such as obesity and lipid fractions using microarray genotyping¹²⁻¹⁶ or sequencing methods.¹⁷⁻²⁰ However, the few previous studies that have systematically addressed differences between genetic architecture of the overall distribution with extremes for complex traits have been small,²¹⁻²³ and hence, it remains largely unknown whether genetic loci affecting the extremes are generalizable to the general population.

Studies of extremely obese individuals have reported thirteen loci at or near genome-wide significance ($P < 5 \times 10^{-7}$),^{14-16,22-26} but not all have shown evidence of association with BMI in the general population.^{4,27} For example, variants in *PCSK1* (rs6232) and *PTER* have been convincingly associated with severe obesity,^{14,25} but have at best shown nominal evidence of association with BMI in large-scale meta-analyses.^{4,28} Although it is possible that other genetic or environmental factors modify the manifestation of these variants producing an extreme phenotype only in selected individuals, it is also conceivable that the extremes are, at least in part, etiologically distinct. Within the extremes of the distribution, there may be etiologically discrete subgroups or enrichment for less common causal variants.¹⁹ Although analyzing the full

distribution is generally more powerful, in cases where there is heterogeneity, analyzing extremes by case-control design may offer superior power.²⁹

The extremes for anthropometric traits, particularly BMI, have been defined in numerous ways, including using tails of the full population distribution (e.g. >95th or >97th percentile) and absolute cutpoints (e.g. $\geq 40 \text{ kg/m}^2$) based on clinical or standard references, and some studies have used a combination of definitions for their discovery and replication. The common denominator for studies addressing 'extremes' (herein used as a more generic term) is that they have dichotomized the trait distribution and analyzed data using a case-control design. Studies suggest that the percentile cutpoint choice and ascertainment strategy utilized may impact the observed risk and subsequent power;^{30,31} however, the consequences of these extreme definitions on discovery and characterization of loci for complex traits have not been systematically evaluated. In the present study, we have used the term 'tails' to describe analyses comparing the upper and lower 5th percentiles of the trait distributions; 'clinical classes of obesity' to describe analyses where controls were subjects with BMI <25 kg/m² and cases were defined as BMI $\geq 25 \text{ kg/m}^2$ for overweight, BMI $\geq 30 \text{ kg/m}^2$ for obesity class I, BMI $\geq 35 \text{ kg/m}^2$ for obesity class II, and BMI $\geq 40 \text{ kg/m}^2$ for obesity class III³²; and 'extremely obese' to describe studies using different sampling designs for selecting their extremely obese cases and controls.

The overall aim of the present study was to use and compare different distribution cutoffs for identification of genetic loci of anthropometric traits. The two specific aims were: 1) to systematically compare findings using these cutoffs with those from the full population distribution, as well as with studies utilizing a different ascertainment strategy; and 2) to draw inferences about the value of these different approaches for sampling within a population-based study. Our focus was primarily on BMI, which is a major risk factor for multiple chronic diseases and of important public health significance,³³ but we also examined height and waist-hip ratio adjusted for BMI (WHR; as a measure of body fat distribution) to verify if our findings could be generalized to other traits. To address these aims, we performed a genome-wide search for genetic determinants of the tails (defined as the upper vs. lower 5th percentile of the trait distribution) of BMI, height and WHR and for comparison, clinical classes of obesity drawn

from populations within the GIANT (Genetic Investigation of ANthropometric Traits) consortium. Association analyses were conducted in a study base (or sampling frame) of up to 168,267 individuals with follow-up of the 273 most significantly associated loci in a study base of up to 109,703 additional individuals. Further, systematic comparisons were conducted to assess differences in genetic inheritance and distribution of risk variants between the extremes and general population for these anthropometric traits.

Results

To first evaluate the contribution of common SNPs to the tails and clinical classes of obesity and discover new loci, we conducted meta-analyses of GWAS of six obesity-related traits (tails of BMI and WHR, overweight, obesity class I, II and III), as well as tails of height, utilizing results for ~2.8 million genotyped or imputed SNPs. Stage 1 analyses included 51 studies with study bases of 158,864 (BMI), 168,267 (height) and 100,605 (WHR) individuals of European ancestry (see **Supplementary Table 1** for number of cases and controls per phenotype; **Supplementary 2-5** for study characteristics). We observed an enrichment of SNPs with small P-values compared to the null distribution for all seven traits (Q-Q plots, **Supplementary Fig. 1-2**). The excess was diminished after exclusion of loci previously established for the overall distributions or extremes of these traits, but some enrichment remained, especially for tails of height and to a lesser extent for overweight, obesity class I and II. In total, 69 loci (defined as separated by at least 1 Mb) were associated at $P < 5 \times 10^{-8}$ with at least one trait (**Supplementary Fig. 3-4**).

To identify and validate loci for these traits, SNPs for which associations reached $P < 5 \times 10^{-6}$ in the stage 1 analyses were taken forward for follow-up (stage 2) in 12 studies with *in silico* GWAS data and 24 studies with MetaboChip data with study bases of 109,703 (BMI), 107,740 (height) and 75,220 (WHR) (**Supplementary Tables 1-5**).

BMI-Related Traits

Seventeen SNPs were taken forward to stage 2 in up to 4,900 and 4,891 individuals from the upper and lower tails of BMI, respectively. Ten SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$) in the joint meta-analysis of stage 1 and stage 2, but all had been previously

identified as loci associated with BMI in the general population.⁴ A total of 118 SNPs were included in stage 2 for clinical classes of obesity, which included up to 1,162 cases and 22,307 controls for obesity class III, and 65,332 cases and 39,294 controls for overweight. Of the 62 SNPs that showed $P < 5 \times 10^{-8}$ in the joint meta-analyses for at least one obesity class (**Supplementary Table 6**), seven were novel, explaining an additional 0.09% of the variability in BMI (**Supplementary Table 7**). These included one locus for overweight (*RPTOR*), three loci for obesity class I (*GNAT2*, *MRPS33P4*, *ADCY9*), two loci for obesity class II (*HS6ST3*, *ZZZ3*), and one locus associated with both overweight and obesity class I (*HNF4G*) (**Table 1, Supplementary Fig. 5-7**). Although these loci were identified for specific clinical classes of obesity, all novel loci showed consistent effect direction across the tails of BMI and the other class of obesity, and most *P*-values were significant ($P < 0.007$, Bonferroni-corrected for 7 SNPs), except for obesity class III and the tails of BMI (presumably due to lower statistical power for these traits; **Table 2**).

Among the novel obesity loci, at least four are located near genes of high biological relevance. In particular, rs7503807 for overweight, is located within the regulatory associated protein of the MTOR, complex 1 gene (*RPTOR*), which regulates cell growth in response to nutrient and insulin levels,³⁴ and within 500 kb of the BAI1-associated protein 2 (*BAIAP2*), which encodes a brain-specific angiogenesis inhibitor (BAI1)-binding protein that regulates insulin uptake in the central nervous system. The overweight and obesity class I SNP rs4735692 is located downstream of the hepatocyte nuclear factor 4-gamma gene (*HNF4G*). Mutations in *HNF4A*, a closely related gene that forms a heterodimer with HNF4G to activate gene transcription,³⁵ cause maturity onset diabetes of the young type 1,³⁶ and a common variant near *HNF4A* was found to be associated with type 2 diabetes (T2D) in east Asians.³⁷ The obesity class I SNP rs2531995 is located within adenylate cyclase 9 (*ADCY9*), which catalyzes the formation of cyclic AMP from ATP. This SNP was found to be associated with *ADCY9* expression in several tissue types (**Supplementary Table 8**). Loci near other adenylate cyclase genes have been associated with several T2D-related traits, such as glucose homeostasis and susceptibility to T2D (*ADCY5*).^{38,39} The obesity class II SNP rs17024258 is located 207kb from the lipid-related gene sortilin (*SORT1*), which is expressed in multiple cell types and has been reported to be involved in insulin responsiveness in adipose cells.⁴⁰ Decreased levels of sortilin have been

observed in adipose tissues of morbidly obese humans and mice, and in skeletal muscle of obese mice.⁴¹ A more comprehensive summary of the biological relevance of the genes nearest to all novel loci is given in the **Supplementary Note**.

Tails of Height

A total of 134 SNPs from stage 1 were taken forward to stage 2 in up to 4,872 and 4,831 individuals from the upper and lower tails of height, respectively. Of the 95 SNPs that reached $P < 5 \times 10^{-8}$ in the joint meta-analysis of stage 1 and stage 2 (**Supplementary Table 6**), four novel loci (*IGFBP4*, *H6PD*, *RSRC1*, *PPP2R2A*) were identified for tails of height (**Table 1**, **Supplementary Fig. 8**). The contribution of the four loci to the overall height variability was $\leq 0.02\%$ (**Supplementary Table 7**).

Two of the novel loci are located near genes that seem particularly relevant to height. rs584438 is located approximately 500 bp upstream of *IGFBP4*, which codes for insulin-like growth factor binding protein 4, and is in linkage disequilibrium ($r^2=0.87$) with another SNP (rs598892) that results in a synonymous amino acid change in *IGFBP4*. *IGFBP4* binds to IGF1 and IGF2,⁴² which have an important role in childhood growth. In blood, this same SNP showed a significant association with the expression of *TNS4* (**Supplementary Table 8**), which interacts with beta-catenin,⁴³ a critical component of the canonical Wnt pathway related to bone formation.⁴⁴ The height SNP rs2362965 lies 285 kb from *SHOX2*, a homolog to the X-linked, pseudoautosomal *SHOX* (short stature homeobox) gene family, which plays a major role in skeletal limb development.

Tails of Waist-Hip Ratio

Ten SNPs were taken forward to stage 2 in 3,351 and 3,352 individuals from the upper and lower tails of WHR, respectively. The four SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$; **Supplementary Table 6**) have been previously identified as WHR loci in the general population.⁷

Comparisons of novel and known loci on the tails, obesity classes, and full distribution

We assessed the impact of our novel loci on the full distribution of these anthropometric traits using data from studies included in stage 1 and stage 2. In the full distribution, evidence of association ($P < 0.005$, Bonferroni-corrected for 11 SNPs) with consistent effect direction was observed with BMI for all novel obesity-related trait loci and with height for all novel loci identified for tails of height (**Table 2**). None of the loci were associated with WHR, suggesting that these obesity loci are primarily associated with overall adiposity, rather than with fat distribution.

Within GIANT, we previously identified 32 loci associated with BMI.⁴ There is considerable overlap of samples with the current study, so it is not unexpected that we observed that the effects of all established BMI loci were directionally consistent between the prior study of overall BMI and the obesity-related traits in the present study (**Supplementary Table 9**). Twenty-seven out of 32 SNPs were significantly associated with the tails of BMI ($P < 0.0016$, Bonferroni-corrected). Although only half of the SNPs were significantly associated with obesity class 3, presumably due to the smaller sample size and reduced power, the majority of SNPs were significantly associated with obesity class 2 and all with obesity class 1 and overweight.

Impact of ascertainment strategy on discovered and known loci

Effect of our novel loci in other studies of extremely obese

Both empirical¹⁶ and theoretical work²⁹ has shown that genetic architecture may differ, the more extreme the selection, suggesting that the ascertainment strategy may impact observed results.³¹ To evaluate impact of ascertainment strategy, we also performed *in silico* look-ups of all SNPs we found to be associated with BMI-related traits in five studies that applied other ascertainment strategies for defining extremely obese (**Supplementary Tables 2-5, bottom panel**; total $n_{\text{cases}}=6,848$; $n_{\text{controls}}=7,023$). Four studies recruited participants from specialized clinics or hospitals based on absolute or percentile-derived cutoffs, and one study utilized liability-based (women) and standard-based (men) percentile cut-points. We performed a meta-analysis of these five studies and observed directionally consistent associations for all BMI-associated SNPs (**Supplementary Table 10**). The effect sizes in these extreme obesity

studies were similar to those observed for tails of BMI in our analysis ($P_{\text{heterogeneity}} > 0.007$ for all SNPs, Bonferroni-corrected). Four out of seven novel obesity-related loci displayed significance at $P < 0.007$ (Bonferroni-corrected) in these extremely obese studies.

Effect of loci previously identified in extremely obese samples in our study

Previous studies of extreme childhood and/or adult obesity using different ascertainment strategies have reported genome-wide significant or near genome-wide significant associations ($P < 5 \times 10^{-7}$) with *FTO*, *MC4R*, *TMEM18*, *FAIM2*, *TNKS*, *HOXB5*, *OLFM4*, *NPC1*, *MAF*, *PTER*, *SDCCAG8*, *PCSK1* (rs6235 and rs6232) and *KCNMA1*.^{14-16,22-26} With the exception of *PCSK1* (rs6232) for tails of BMI and *MAF* for tails of BMI and obesity class II, all associations showed consistent directions of effect across the BMI-related outcomes (**Supplementary Table 11**). Of the 13 loci, replication at a significance level of $P < 0.004$ (Bonferroni-corrected) was observed for four SNPs (*FTO*, *MC4R*, *TMEM18*, *FAIM2*) for the tails of BMI and all clinical classes of obesity. Two loci, *MAF* and *KCNMA1*, which have thus far only been reported for extreme obesity, were not significantly associated with any of our traits at either a Bonferroni-corrected or nominal significance threshold ($P < 0.05$).

Empirical power comparison of the extremes and full distribution

If the extremes have different genetic inheritance or are etiologically more homogenous than the full distribution, analyzing extremes or tails of the distribution by case-control design may offer superior power. To test this empirically, we conducted meta-analyses of the full distributions of BMI and height with all studies included in stage 1 and stage 2. Only two (*IGFBP4* and *H6PD*) out of four novel loci for tails of height reached genome-wide significance ($P < 5 \times 10^{-8}$) using the full height distribution (**Table 2**). Four (*GNAT2*, *ZZZ3*, *HNF4G*, and *RPTOR*) out of seven novel loci identified for clinical classes of obesity achieved genome-wide significance for the full BMI distribution. The remaining loci had P -values $< 5 \times 10^{-5}$ in the full distribution and thus, would likely have been detected with a larger sample size.

Systematic comparisons of the genetic inheritance and distribution of SNPs between the tails and full distribution

To investigate differences in genetic architecture between the tails and full distributions, we estimated whether the observed genetic effects in tails of BMI, height and WHR were different from what would be expected based on the full distributions of corresponding traits. To do this, we first estimated the expected effect for each SNP in the tails based on the full distribution in each study and then meta-analyzed the expected associations across studies. The Q-Q plots of *P*-values testing differences between the observed and expected (**Fig. 1** and **Supplementary Fig. 9**) did not show any enrichment, indicating that effect sizes observed in tails and those expected based on the overall distribution were similar. Further, comparable results were observed for the 32 SNPs previously associated with BMI in Speliotes *et al*⁴, as well as for previously published and novel extreme obesity loci (**Supplementary Table 12**).

To further compare genetic inheritance of the tails with the full distribution, we used a 'polygene approach'.⁴⁵ The meta-analysis results of tails and full distribution were used to create two polygenetic scores (by summing the number of risk alleles at each SNP) in six studies (**Supplementary Table 13**). We found that the polygene score based on the full BMI distribution consistently explained more of the variance than the score based on the tails (e.g. 15.3% vs. 6.4% at $P < 0.05$) (**Fig. 2, Supplementary Table 14**). Similar results were observed for height and WHR (**Supplementary Fig. 10**). On liability scale, the variance explained by the two polygene scores was similar for different BMI-related outcomes (**Supplementary Fig. 11**) and different percentile cutpoints used to define the tails (data not shown), suggesting that the fraction of the overall variance explained by SNPs is not influenced by the outcome categorization, but by the ability to accurately rank and estimate the beta coefficients of the association, which is better achieved by using the entire study population instead of the tails. Our results also indicate that genetic determinants for the tails are similar to those for the full distribution and that common variant loci contribute to extreme phenotypes. However, it should be noted that our analyses of the upper and lower 5 percentiles of the distribution (tails) does not necessarily extend to more extreme cut-offs, such as the top and bottom 1st percentiles.

Allelic heterogeneity at known and discovered loci

To explore enrichment for allelic heterogeneity in the tails and clinical classes of obesity, we performed conditional analyses using a method recently described by Yang *et al.*⁴⁶ In these analyses, we found secondary signals that reached genome-wide significance ($P < 5 \times 10^{-8}$) at 17 loci, including one locus for tails of BMI (*FTO*), 13 loci for tails of height (*PTCH1* [two signals], *GHSR*, *EDEM2*, *C6orf106*, *CRADD*, *EFEMP1*, *HHIP*, *FBXW11*, *NPR3*, *C2orf52*, *BCKDHB*, *EFR3B*), one for tails of WHR (*RSPO3*), two for overweight (*MC4R*, *FANCL*), and one for obesity class I (*FANCL*; **Supplementary Table 15**). Whereas the secondary signals for tails of BMI (*FTO*) and WHR (*RSPO3*), and overweight and obesity class I (*FANCL*) have not been established previously, all 13 height loci identified here, as well as the *MC4R* locus have previously been shown to have allelic heterogeneity in the general population,^{7,9} suggesting that there is no enrichment in the tails for secondary signals (**Supplementary Fig. 12-14**).

We also looked for evidence of enrichment of unobserved low-frequency variants by conducting haplotype analyses within known and novel loci, since haplotypes constructed from common SNPs may tag low-frequency variants that are enriched in the tails of the trait distributions, but are rarer in the general population. Using genotype data from the largest studies, three signals of association were observed for tails of height that exceeded conservative prior odds of association of one in 30,000: *ID4* (Bayes factor: 118,839), *LIN28B* (Bayes factor: 105,478) and *DLEU7* (Bayes factor: 66,599) (**Supplementary Table 16**). However, for all three loci, association signals were characterized by two clusters of haplotypes (both common and rare) and were not consistent with an enrichment of unobserved low-frequency causal variants in the distribution tails.

Discussion

In our meta-analysis of genome-wide association studies of up to 263,407 individuals of European ancestry, we identified 165 loci associated with tails (upper vs. lower 5th percentile) of BMI, height, and WHR and/or clinical classes of obesity. Eleven of these loci have not previously been associated with anthropometric traits. Several of the novel loci were located near strong biological candidate genes, such as *IGFBP4* and *SHOX2* for tails of height, and

HNF4G and *ADCY9* for overweight/obesity class I, suggesting future areas of research. Although by using different distribution cutoffs we discovered additional loci that would not have been identified as genome-wide significant using the full distribution of the same study samples, there is no evidence to suggest that the clinical classes of obesity are etiologically distinct, and the majority of evidence indicates that the extremes share many of the same loci with the general population.

To assess the impact of different distribution cutpoints on genetic variants associated with the extremes, we chose to evaluate the 5% tails of the distribution and clinical classes of obesity, specifically obesity classes II and III. Although others have ascertained extremes differently, all variants associated with obesity-related traits in our meta-analysis were found to have directionally consistent results in five independent studies of extremely obese samples. Of the 13 loci previously identified as associated with extreme obesity,^{14-16,22-26} nearly all (except *PCSK1* rs6232 and *MAF*) showed a consistent direction of effect for the tails of BMI. Only two loci (*MAF* and *KCNMA1*), originally identified for early-onset and morbid adult obesity,^{14,26} failed to replicate for any of our BMI-related outcomes. While it is possible that we had insufficient power if there was a substantial winner's curse present in the initial publications, it is also conceivable that these susceptibility loci are population-specific, only contribute to risk at younger ages,⁴⁷ represent false positive findings, or tag rare causal variants that are difficult to detect in population-based samples.

Since our study was based on GWAS data, we were not well suited to address the role of rare variants in extreme traits. Although the haplotype-based analyses revealed strong associations of haplotypes in three genes with tails of height, which could suggest that they are tagged by rare variants, such putative variants could not be established using our approach. The suggestion that rare variants could be more important in extremes of complex traits needs to be addressed using other designs, such as resequencing projects or using the new Exome Chip microarrays that are currently being analyzed in many large study samples.

Our systematic comparisons between extremes and full distribution yielded several important insights that also may be informative for other complex traits. When comparing observed genetic effects in tails with expected effects extrapolated from overall distributions of

corresponding traits, we did not observe any systematic differences. Further, we showed that the polygene score based on the full distribution explained a larger proportion of variance than the score based on the tails. Taken together with the finding that half of our novel loci were associated at genome-wide significant level in the overall distribution, this implies that there is limited etiologic heterogeneity in these anthropometric traits. Our analysis shows that while some common variants can have larger effects in the extremes, these effects as a whole are not larger than expected based on the effects in the overall distribution. Further, while rare variants specific to the extremes may still exist, the extremes share most of the common loci with the overall distribution.

Conclusions that can be drawn from these observations are that when having access to data for the full distribution, case-control analyses using extremes can be useful to find additional loci. Although the analyzing the full distribution is generally more powerful, small amounts of heterogeneity in the distribution may allow for the identification of additional loci by analyzing the data using different cut-points, such as the tails. Further, as in most cases when resources are limited, our results indicate that a strategy with selection of individuals from the extremes for genetic analyses could be a cost-effective approach and will likely yield loci that are relevant and largely generalizable to the full population. Compatible with those of recent, smaller studies,²¹⁻²³ our results show convincingly that this theoretically appealing approach also holds empirically.

In conclusion, in our large GWAS meta-analysis including up to 263,407 individuals, we identified four new loci influencing height detected at the tails, as well as seven new loci for clinical classes of obesity. Consistent with theoretical predictions and previous smaller studies, our results show that there is a large overlap in terms of genetic structure and distribution of variants between traits based on different distribution cutoffs with those from population-level studies, but additional insight may still be gained from evaluating the extremes. Our results are informative for designing future genetic studies of obesity as well as other complex traits.

Acknowledgments

A full list of acknowledgments appears in the Supplementary Note.

Aarno Koskelo Foundation; Academy of Finland; Agency for Science, Technology and Research of Singapore; Australian National Health and Medical Research Council; Australian Research Council; BDA Research; BioSHaRE Consortium; British Heart Foundation; Cedars-Sinai Board of Governors' Chair in Medical Genetics; Centre for Clinical Research at the University of Leipzig; Centre of Excellence in Genomics and University of Tartu; Chief Scientist Office of the Scottish Government; City of Kuopio and Social Insurance Institution of Finland; Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano; Donald W. Reynolds Foundation; Dutch Ministry for Health, Welfare and Sports; Dutch Ministry of Education, Culture and Science; Dutch BBRMI-NL; Dutch Brain Foundation; Dutch Centre for Medical Systems Biology; Dutch Diabetes Research Foundation; Dutch Government Economic Structure Enhancing Fund; Dutch Inter University Cardiology Institute; Dutch Kidney Foundation; Dutch Ministry of Economic Affairs; Dutch Ministry of Justice; Dutch Research Institute for Diseases in the Elderly; Eleanor Nichols endowments; Emil Aaltonen Foundation; Erasmus Medical Center and Erasmus University; Estonian Government; European Commission; European Regional Development Fund; European Research Council; European Science Foundation; Faculty of Biology and Medicine of Lausanne; Finland's Slot Machine Association; Finnish Cultural Foundation; Finnish Diabetes Research Foundation; Finnish Foundation for Cardiovascular Research; Finnish Funding Agency for Technology and Innovation; Finnish Heart Association; Finnish Medical Society; Finnish Ministry of Education and Culture; Finnish Ministry of Health and Social Affairs; Finnish National Institute for Health and Welfare; Finnish Social Insurance Institution; Finska Läkaresällskapet; Folkhälsan Research Foundation; Foundation for Life and Health in Finland; French Ministry of Research; French National Research Agency; Genetic Association Information Network; German Diabetes Association; German Federal Ministry of Education and Research; German Ministry of Cultural Affairs; German National Genome Research Network; German Research Foundation; GlaxoSmithKline; Göteborg Medical Society; Greek General Secretary of Research and Technology; Gyllenberg Foundation; Health Care Centers in Vasa, Närpes and Korsholm; Heinz Nixdorf Foundation; Helmholtz Zentrum

München, German Research Center for Environmental Health; Icelandic Heart Association; Icelandic Parliament; Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH; Italian Ministry of Education, Universities and Research; Italian Ministry of Health; Juho Vainio Foundation; Juvenile Diabetes Research Foundation International; Knut and Alice Wallenberg Foundation; Kuopio, Tampere and Turku University Hospital Medical Funds; Leducq Foundation; Lundberg Foundation; March of Dimes; Munich Center of Health Sciences as part of LMUinnovativ; Municipal Health Care Center and Hospital in Jakobstad; Municipality of Rotterdam; Närpes Health Care Foundation; National Alliance for Research on Schizophrenia and Depression Young Investigator Awards; Netherlands Genomics Initiative; Netherlands organization for health research and development; NHSBT; National Institute of Health; Nordic Center of Cardiovascular Research; Nordic Center of Excellence in Disease Genetics; Nordic Centre of Excellence on Systems biology in controlled dietary interventions and cohort studies; Northern Netherlands Collaboration of Provinces; Novo Nordisk Foundation; Ollqvist Foundation; Orion-Farmos Research Foundation; Paavo Nurmi Foundation; Päivikki and Sakari Sohlberg Foundation; Perklén Foundation; Petrus and Augusta Hedlunds Foundation; Province of Groningen; Republic of Croatia Ministry of Science, Education and Sport; Reynold's Foundation; Royal Society; Samfundet Folkhälsan; Signe and Ane Gyllenberg foundation; Sigrid Juselius Foundation; Social Ministry of the Federal State of Mecklenburg-West Pomerania; Sophia Foundation for Medical Research; South Tyrolean Sparkasse Foundation; Southern California Diabetes Endocrinology Research Center; Stockholm County Council; Strategic Cardiovascular Program of Karolinska Institutet; Strategic support for epidemiological research at Karolinska Institutet; Susan G. Komen Breast Cancer Foundation; Swedish Ministry for Higher Education; Swedish Cancer Society; Swedish Cultural Foundation in Finland; Swedish Diabetes Association; Swedish Foundation for Strategic Research; Swedish Heart-Lung Foundation; Swedish Medical Research Council; Swedish Ministry of Education; Swedish Research Council; Swedish Royal Academy of Science; Swedish Society for Medical Research; Swedish Society of Medicine; Swiss National Science Foundation; Tampere Tuberculosis Foundation; The Great Wine Estates of the Margaret River region of Western Australia; The Paul Michael Donovan Charitable Foundation; Torsten and Ragnar Söderberg

Foundation; UK Cancer Research; UK Diabetes Association; UK Heart Foundation; UK Medical Research Council; UK National Institute for Health Research, Biomedical Research Centre; UK West Anglia Primary and Community Care; University Medical Center Groningen and University of Groningen; Västra Götaland Foundation; VU University: Institute for Health and Care Research and Neuroscience Campus Amsterdam; Wellcome Trust; Yrjö Jahnsson Foundation.

Author contributions

Steering committee (oversaw the consortium)

Goncalo R Abecasis, Themistocles Assimes, Ines Barroso, Sonja Berndt, Michael Boehnke, Ingrid Borecki, Panos Deloukas, Caroline Fox, Tim Frayling, Leif Groop, Talin Haritunian, Iris Heid, David Hunter, Erik Ingelsson, Robert C Kaplan, Ruth JF Loos, Mark McCarthy, Karen Mohlke, Kari E North, Jeffrey R O'Connell, David Schlessinger, David Strachan, Unnur Thorsteinsdottir, Cornelia van Duijn

Writing group (drafted and edited manuscript)

Sonja I Berndt, Mary F Feitosa, Andrea Ganna, Stefan Gustafsson, Erik Ingelsson, Anne E Justice, Cecilia M Lindgren, Ruth JF Loos, Reedik Mägi, Mark McCarthy, David Meyre, Keri L Monda, Andrew P Morris, Kari E North, André Scherag, Elizabeth K Speliotes, Eleanor Wheeler, Cristen J Willer

Data cleaning and preparation

Sonja I Berndt, Damien C Croteau-Chonka, Felix R Day, Tõnu Esko, Tove Fall, Teresa Ferreira, Stefan Gustafsson, Iris Heid, Erik Ingelsson, Anne U Jackson, Hana Lango Allen, Cecilia M Lindgren, Jian'an Luan, Reedik Mägi, Joshua C Randall, André Scherag, Elizabeth K Speliotes, Gudmar Thorleifsson, Sailaja Vedantam, Thomas W Winkler, Andrew R Wood

Statistical Advisors

Sang Hong Lee, Benjamin M Neale, Yudi Pawitan, Peter M Visscher, Jian Yang, Dan-Yu Lin, Yi-Juan Hu

Gene-expression (eQTL) analyses

Liming Liang, William O Cookson, Miriam F Moffatt, Goncalo R Abecasis, Valgerdur Steinthorsdottir, Gudmar Thorleifsson, Josine L Min, George Nicholson, Fredrik Karpe, Mark I McCarthy, Eric E Schadt

Project design, management and coordination of contributing studies

Stage 1 – Genome-wide association studies

(ADVANCE) Themistocles L Assimes, Carlos Iribarren; (AGES) Vilmundur Gudnason, Tamara B Harris, Lenore J Launer; (ARIC) Eric Boerwinkle, Kari E North; (B58C) David P Strachan; (BRIGHT study) Mark J Caulfield, Patricia B Munroe; (CAPS) Erik Ingelsson; (CHS) Barbara McKnight, Bruce M Psaty; (CoLaus) Vincent Mooser, Peter Vollenweider, Gérard Waeber; (COROGENE) Markku S Nieminen, Juha Sinisalo; (deCODE) Kari Stefansson, Unnur Thorsteinsdottir; (DGI) Leif C Groop, Joel N Hirschhorn; (EGCUT) Andres Metspalu; (EPIC) Kay-Tee Khaw, Ruth JF Loos, Nicholas J Wareham; (ERF) Ben A Oostra, Cornelia M van Duijn; (FamHS) Ingrid B Borecki, Michael A Province; (Fenland) Ruth JF Loos, Nicholas J Wareham; (FRAM) L A Cupples, Caroline S Fox; (FUSION GWAS) Michael Boehnke, Karen L Mohlke; (Genmets) Antti Jula, Samuli Ripatti, Veikko Salomaa; (GerMIFS1) Jeanette Erdmann, Heribert Schunkert; (GerMIFS2) Christian Hengstenberg, Klaus Stark; (GOOD) Claes Ohlsson; (HBCS) Johan G Eriksson; (KORA S3) H.- E Wichmann; (KORA S4) Christian Gieger, Thomas Illig, Wolfgang Koenig, Annette Peters; (MGS) Pablo V Gejman, Douglas F Levinson; (MICROS (SOUTH TYROL)) Peter P Pramstaller; (MIGEN) Joel N Hirschhorn, Sekar Kathiresan; (NESDA) Brenda Penninx; (NFBC 1966) Marjo-Riitta Jarvelin; (NHS) Lu Qi; (Nijmegen Biomedical Study) Lambertus A Kiemeneij; (NSPHS) Ulf Gyllensten; (NTR) Dorret I Boomsma; (ORCADES) James F Wilson, Alan F Wright; (PLCO) Stephen J Chanock, Sonja I Berndt; (PROCARDIS) Martin Farrall, Hugh Watkins; (RS-I) Fernando Rivadeneira, André G Uitterlinden; (RUNMC) Lambertus A. Kiemeneij; (SardiNIA) David Schlessinger; (SASBAC) Erik Ingelsson; (SHIP) Henri Wallaschofski; (Sorbs) Michael Stumvoll, Anke Tönjes; (TwinsUK) Tim D Spector; (VIS) Igor Rudan; (WGHS) Paul M Ridker; (WTCC-T2D)

Mark I McCarthy; (WTCCC-CAD) Anthony J Balmforth, Alistair S Hall, Nilesh J Samani; (YFS) Mika Kähönen, Terho Lehtimäki, Olli Raitakari, Jorma Viikari

Stage 2 – MetaboChip/in silico replication

(AMC-PAS) Kees G Hovingh; (B58C) Chris Power; (BHS) Lyle J Palmer; (DILGOM) Kari Kuulasmaa, Veikko Salomaa; (DPS) Matti Uusitupa; (DR's EXTRA) Timo A Lakka, Rainer Rauramaa; (EPIC, Fenland and Ely) Claudia Langenberg, Ruth JF Loos, Nicholas J Wareham; (FIN-D2D 2007) Sirkka M Keinänen-Kiukaanniemi, Timo E Saaristo; (GLACIER) Paul W Franks; (Go-DARTS (Dundee)) Andrew D Morris, Colin NA Palmer; (HNR) Karl-Heinz Jöckel; (HUNT 2) Kristian Hveem; (Hypergenes) Daniele Cusi; (IMPROVE) Ulf de Faire, Anders Hamsten, Elena Tremoli; (KORA S3) Iris Heid; (LifeLines Cohort Study) Harold Snieder, Melanie M Van der Klauw, Bruce HR Wolffenbuttel; (LURIC) Bernhard O Boehm, Winfried März, Bernhard R Winkelmann; (METSIM) Johanna Kuusisto, Markku Laakso; (MORGAM) Philippe Amouyel, Paolo Brambilla, Marco M Ferrario, Jean Ferrières, Frank Kee, David-Alexandre Tregouet, Jarmo Virtamo; (NSHD) Diana Kuh; (PIVUS) Erik Ingelsson; (PLCO2) Sonja I Berndt, Stephen J Chanock; (PREVEND) Pim van der Harst; (QIMR) Nicholas G Martin, Grant W Montgomery, Andrew Heath, Pamela Madden; (RS-II) Albert Hofman, Joyce BJ van Meurs; (RS-III) Cornelia M Van Duijn, Jacqueline CM Witteman; (Swedish Twin Reg.) Erik Ingelsson; (THISEAS / AMCPAS / CARDIOGENICS) Panos Deloukas; (THISEAS) George V Dedoussis; (TRAILS) Albertine J Oldehinkel; (Tromsø 4) Inger Njølstad; (TWINGENE) Erik Ingelsson; (ULSAM) Erik Ingelsson; (Whitehall II) Aroon Hingorani, Mika Kivimäki; (WTCC-T2D) Mark I McCarthy, Cecilia M Lindgren

Other contributing studies: clinically extremes

(French Extreme Obesity Study) David Meyre, Philippe Froguel; (GEO-IT) Anna Maria Di Blasio; (Essen Obesity Study, Essen Case-Control & Essen Obesity Trio GWAS) Johannes Hebebrand, Anke Hinney; (GOYA) Thorkild IA Sørensen, Ellen A Nohr

Genotyping of contributing studies

Stage 1 – Genome-wide association studies

(ADVANCE) Devin Absher; (ARIC) Eric Boerwinkle; (B58C) Wendy L McArdle; (CAPS) Henrik Grönberg; (CHS) Talin Haritunians; (CoLaus) Vincent Mooser; (COROGENE) Markus Perola; (EGCUT) Tõnu Esko, Lili Milani; (EPIC) Inês Barroso; (ERF) Ben A Oostra, Cornelia M van Duijn; (FamHS) Ingrid B Borecki, Michael A Province, Aldi T Kraja; (Fenland) Jian'an Luan; (Genmets) Samuli Ripatti; (GOOD) Claes Ohlsson, John-Olov Jansson, Mattias Lorentzon; (HBCS) Aarno Palotie, Elisabeh Widén; (MGS) Pablo V Gejman, Alan R Sanders; (MICROS (SOUTH TYROL)) Andrew A Hicks; (NHS) Frank B Hu, David Hunter; (NSPHS) Åsa Johansson; (NTR and NESDA) Jouke- Jan Hottenga; (ORCADES) James F Wilson; (PLCO) Stephen J Chanock, Kevin B Jacobs; (RS-I) Fernando Rivadeneira, André G Uitterlinden, Karol Estrada, Carolina Medina-Gomez; (SardiNIA) Mariano Dei; (SASBAC) Per Hall, Jianjun Liu; (SHIP) Georg Homuth; (TwinsUK) Massimo Mangino, So-Youn Shin, Nicole Soranzo; (VIS) Caroline Hayward, Veronique Vitart; (WGHS) Daniel I Chasman; (WTCC-T2D) Mark I McCarthy; (WTCCC-CAD) Anthony J Balmforth, Alistair S Hall, Nilesh J Samani; (YFS) Terho Lehtimäki

Stage 2 – MetaboChip/in silico replication

(BHS) Lyle J Palmer, John Beilby; (CARDIOGENICS) Sarah Edkins, Sarah E Hunt; (DPS) Amy J Swift; (DR's EXTRA) Mustafa Atalay; (EPIC, Fenland and Ely) Jian'an Luan, Ken K Ong; (FIN-D2D 2007) Peter S Chines; (FUSION) Francis S Collins, Jouko Saramies, Jaakko Tuomilehto; (GLACIER) Inês Barroso, Sarah Edkins; (Go-DARTS (Dundee)) Colin NA Palmer; (HNR) Thomas W Mühleisen; (HUNT 2) Narisu Narisu; (Hypergenes) Francesca Frau; (KORA S3) Harald Grallert; (KORA S4) Thomas Illig; (LifeLines Cohort Study) Harold Snieder, Bruce HR Wolffenbuttel, Marcel Bruinenberg, Lude Franke; (NSHD) Diana Kuh, Ken K Ong, Andrew Wong; (PIVUS) Erik Ingelsson, Lars Lind; (PLCO2) Stephen J Chanock, Kevin B Jacobs, Zhaoming Wang; (PREVEND) Pim van der Harst, Folkert W Asselbergs; (QIMR) Nicholas G Martin, Grant W Montgomery, Andrew C Heath, Pamela A Madden; (RS-II) Marjolein Peters, Mariano Dei; (Swedish Twin Reg.) Erik Ingelsson, Patrik K Magnusson, Nancy Pedersen; (THISEAS / AMCPAS / CARDIOGENICS) Kathleen Stirrups; (TRAILS) Albertine J Oldehinkel, Ilja M Nolte, Jana V Van Vliet-Ostaptchouk; (Tromsø 4) Lori L Bonnycastle; (TWINGENE) Erik Ingelsson, Anders Hamsten, Nancy Pedersen; (ULSAM) Erik Ingelsson; (Whitehall II) Claudia Langenberg; (WTCC-T2D) Mark I McCarthy

Other contributing studies: clinically extremes

(French Extreme Obesity Study) David Meyre, Philippe Froguel; (GEO-IT) Davide Gentilini;
(GOYA) Lavinia Paternoster, David M Evans

Phenotyping of contributing studies

Stage 1 – Genome-wide association studies

(ARIC) Eric Boerwinkle; (B58C) David P Strachan; (BRIGHT study) John M Connell; (CAPS) Henrik Grönberg; (CHS) Bruce M Psaty; (CoLaus) Peter Vollenweider, Gérard Waeber; (COROGENE) Juha Sinisalo, Marja-Liisa Lokki; (EGCUT) Andres Metspalu, Krista Fischer; (EPIC) Ruth JF Loos; (ERF) Ben A Oostra, Cornelia M van Duijn; (FamHS) Ingrid B Borecki, Michael A Province, Mary F Feitosa; (Fenland) Ruth JF Loos; (FRAM) Caroline S Fox; (Genmets) Antti Jula, Veikko Salomaa; (GerMIFS1) Stefan Schreiber; (GerMIFS2) Annette Peters; (GOOD) Claes Ohlsson, John-Olov Jansson, Mattias Lorentzon, Liesbeth Vandenput; (MGS) Pablo V Gejman, Alan R Sanders, Douglas F Levinson; (NFBC 1966) Anna-Liisa Hartikainen, Jaana H Laitinen, Anneli Pouta; (NHS) Lu Qi; (Nijmegen Biomedical Study) Femmie de Vegt, Martin den Heijer, Sita H Vermeulen; (NSPHS) Åsa Johansson, Ulf Gyllensten; (NTR and NESDA) Gonneke Willemsen; (ORCADES) Harry Campbell, Sarah H Wild; (PLCO) Sonja I Berndt; (RS-I) Fernando Rivadeneira, André G Uitterlinden; (SASBAC) Per Hall; (SHIP) Sabine Schipf; (Sorbs) Anke Tönjes; (TwinsUK) Massimo Mangino, Tim D Spector; (VIS) Ozren Polasek; (WTCC-T2D) Mark I McCarthy; (WTCCC-CAD) Anthony J Balmforth, Alistair S Hall, Nilesh J Samani; (YFS) Mika Kähönen, Olli Raitakari, Jorma Viikari

Stage 2 – Metabochip/in silico replication

(AMC-PAS) Hanneke Basart, Mieke D Trip; (B58C) Chris Power, Elina Hyppönen; (BHS) Lyle J Palmer, John Beilby, Arthur W Musk; (DPS) Jaana Lindström; (EPIC, Fenland and Ely) Ruth JF Loos; (GLACIER) Paul W Franks, Dmitry Shungin; (Go-DARTS (Dundee)) Colin NA Palmer, Andrew D Morris; (Hypergenes) Daniele Cusi, Paolo Manunta; (KORA S3) Barbara Thorand; (KORA S4) Annette Peters; (LifeLines Cohort Study) Bruce HR Wolffenbuttel, Melanie M Van der Klauw;

(METSIM) Alena Stančáková, Pablo V Gejman; (NSHD) Diana Kuh; (PIVUS) Erik Ingelsson, Lars Lind; (PLCO2) Sonja I Berndt; (PREVEND) Gerjan Navis; (QIMR) Nicholas G Martin, Andrew Heath, Pamela Madden; (RS-II) M Carola Zillikens; (RS-III) Jacqueline CM Witteman; (Swedish Twin Reg.) Erik Ingelsson, Patrik K Magnusson, Nancy Pedersen; (THISEAS) Maria Dimitriou, Eirini V Theodoraki; (TRAILS) Ronald P Stolk; (TWINGENE) Erik Ingelsson, Nancy Pedersen; (ULSAM) Erik Ingelsson; (Whitehall II) Meena Kumari; (WTCC-T2D) Mark I McCarthy

Other contributing studies: clinically extremes

(Essen Obesity Study, Essen Case-Control GWAS & Essen Obesity Trio GWAS) Johannes Hebebrand, Anke Hinney; (GEO-IT) Antonio Liuzzi, Stefano Signorini; (GOYA) Thorkild IA Sørensen, Ellen A Nohr

Analyses of contributing studies

Stage 1 – Genome-wide association studies

(ADVANCE) Lindsay L Waite; (AGES) Albert V Smith; (ARIC) Kari E North, Anne E Justice, Keri L Monda; (B58C) David P Strachan; (BRIGHT study) Toby Johnson; (CAPS) Erik Ingelsson, Reedik Mägi; (CHS) Barbara McKnight, Guo Li; (CoLaus) Diana Marek; (COROGENE) Markus Perola; (deCODE) Valgerdur Steinthorsdottir, Gudmar Thorleifsson; (DGI) Elizabeth K Speliotes, Sailaja Vedantam; (EGCUT) Krista Fischer, Tõnu Esko, Evelin Mihailov; (EPIC) Jing Hua Zhao; (ERF) Najaf Amin; (FamHS) Mary F Feitosa; (Fenland) Jian'an Luan; (FRAM) L A Cupples, Nancy L Heard-Costa, Julius S Ngwa; (Genmets) Ida Surakka; (GerMIFS1) Michael Preuss; (GerMIFS2) Inke R König; (GOOD) Claes Ohlsson, John-Olov Jansson, Mattias Lorentzon, Liesbeth Vandenput; (HBCS) Niina Eklund; (KORA S3) Claudia Lamina; (KORA S4) Eva Albrecht; (MGS) Douglas F Levinson, Jianxin Shi; (MICROS (SOUTH TYROL)) Jennifer E Huffman, Åsa Johansson; (MIGEN) Elizabeth K Speliotes, Sailaja Vedantam; (NFBC 1966) Anneli Pouta, Reedik Magi, Joshua C Randall; (NHS) Lu Qi, Tsegaselassie Workalemahu; (NSPHS) Åsa Johansson; (NTR and NESDA) Jouke- Jan Hottenga, Marleen H. de Moor; (ORCADES) Åsa Johansson; (PLCO) Kevin B Jacobs, Sonja I Berndt; (PROCARDIS) Martin Farrall, Anuj Goel, John F Peden; (RS-I) Fernando Rivadeneira, Karol Estrada, Carolina Medina-Gomez; (SARDINIA) Jennifer L Bragg-Gresham,

Serena Sanna; (SASBAC) Erik Ingelsson, Reedik Mägi; (SEARCH) Jonathan Tyrer; (SHIP) Alexander Teumer; (Sorbs) Reedik Mägi, Inga Prokopenko; (TwinsUK) Massimo Mangino, Nicole Soranzo; (VIS) Åsa Johansson; (WGHS) Daniel I Chasman; (WTCC-T2D) Cecilia M Lindgren, Reedik Magi, Joshua C Randall; (WTCCC-CAD) Reedik Magi, Joshua C Randall; (WTCCC-NBS (UKBS-CC)) Antony P Attwood, Reedik Magi, Joshua C Randall, Jennifer G Sambrook, Jonathan C Stephens; (YFS) Olli Raitakari, Terho Lehtimäki

Stage 2 – MetaboChip/in silico replication

(B58C) Elina Hyppönen, Teresa Ferreira; (BHS) Gemma Cadby; (DILGOM) Kati Kristiansson; (DPS) Anne U Jackson; (DR's EXTRA) Anne U Jackson; (EPIC, Fenland and Ely) Jian'an Luan, Ken K Ong; (FIN-D2D 2007) Anne U Jackson; (FUSION) Anne U Jackson; (GLACIER) Paul W Franks, Dmitry Shungin; (HNR) Sonali Pechlivanis, Carolin Pütter; (HUNT 2) Anne U Jackson; (Hypergenes) Francesca Frau, Zoltán Kutalik; (IMPROVE) Rona J Strawbridge; (KORA S3) Iris Heid, Thomas W Winkler; (KORA S4) Martina Müller-Nurasyid; (LifeLines Cohort Study) Marcel Bruinenberg, Lude Franke; (LURIC) Marcus E Kleber; (METSIM) Anne U Jackson; (NSHD) Andrew Wong, Jian'an Luan; (PIVUS) Erik Ingelsson, Stefan Gustafsson; (PLCO2) Sonja I Berndt, Zhaoming Wang; (PREVEND) Pim van der Harst, Irene Mateo Leach; (QIMR) Sarah E Medland, Jian Yang; (RS-II) Marjolein Peters; (Swedish Twin Reg.) Erik Ingelsson, Stefan Gustafsson; (THISEAS / AMCPAS / CARDIOGENICS) Stavroula Kanoni; (TRAILS) Ilja M Nolte, Jana V Van Vliet-Ostaptchouk; (Tromsø 4) Anne U Jackson; (TWINGENE) Erik Ingelsson, Stefan Gustafsson; (ULSAM) Erik Ingelsson, Andrea Ganna, Stefan Gustafsson; (WTCC-T2D) Reedik Magi, Teresa Ferreira

Other contributing studies: clinically extremes

(French Extreme Obesity Study) David Meyre, Cecile Lecoeur, Boris Skrobek; (GEO-IT) Anna Maria Di Blasio, Davide Gentilini; (Essen Obesity Study, Essen Case-Control GWAS & Essen Obesity Trio GWAS) André Scherag, Ivonne Jarick; (GOYA) Lavinia Paternoster, David M Evans

Competing interest statements

Kari Stefansson, Valgerdur Steinthorsdottir, Gudmar Thorleifsson and Unnur Thorsteinsdottir are employed by deCODE Genetics with stock/stock options in the company. Inês Barroso and spouse own stock in Incyte Ltd and GlaxoSmithKline. Eric E. Schadt is Chief Scientific Officer, Pacific Biosciences. Thorkild IA Sørensen has collaborated with Nestlé Research Centre and DSM, The Netherlands, on metabolomics and genomics of obesity and related traits, respectively.

Tables

Table 1. Novel loci reaching genome-wide significance ($P < 5 \times 10^{-8}$) for the tails of anthropometric traits and clinical classes of obesity

SNP	Chr	Position	Nearby gene	Effect allele	Other allele	Effect allele freq	Stage 1				Stage 2*				Stage 1 + Stage 2			
							No. cases	No. controls	OR	P	No. cases	No. controls	OR	P	No. cases	No. controls	OR	P
Height tails																		
rs584438	17	35852698	<i>IGFBP4</i>	c	a	0.617	7830	7850	1.18	1.11E-09	1814	1814	1.19	0.001	9644	9664	1.18	5.22E-12
rs6662509	1	9240191	<i>H6PD</i>	t	c	0.1463	5462	5461	1.23	2.21E-06	3615	3566	1.23	3.37E-05	9077	9027	1.23	3.19E-10
rs2362965	3	159592073	<i>RSRC1/SHOX2</i>	t	a	0.50	7989	7993	1.14	1.45E-07	4819	4775	1.10	0.002	12808	12768	1.12	2.14E-09
rs1594829	8	26261994	<i>PPP2R2A</i>	c	t	0.7688	6693	6697	1.18	5.51E-07	4166	4115	1.11	0.01	10859	10812	1.15	3.88E-08
Obesity Class 2																		
rs7989336	13	95815549	<i>HS6ST3</i>	a	g	0.4704	9825	62114	1.12	5.88E-09	1664	17113	1.04	0.25	11489	79226	1.10	1.06E-08
rs17381664	1	77820919	<i>ZZZ3</i>	c	t	0.3923	9833	62114	1.11	7.61E-08	5351	33841	1.05	0.04	15184	95955	1.09	2.85E-08
Obesity Class 1																		
rs17024258	1	109948844	<i>GNAT2</i>	t	c	0.0364	18662	38427	1.23	1.41E-06	8956	15471	1.28	1.12E-06	27618	53898	1.25	8.66E-12
rs4735692	8	76778218	<i>HNF4G</i>	a	g	0.5834	32675	65697	1.07	5.03E-08	22086	38352	1.04	0.005	54761	104049	1.06	2.48E-09
rs13041126	20	50526403	<i>MRPS33P4</i>	t	c	0.7179	32020	64015	1.07	3.05E-07	22088	37595	1.04	0.007	54108	101610	1.06	2.16E-08
rs2531995	16	3953468	<i>ADCY9</i>	t	c	0.6146	32433	65542	1.06	3.17E-06	6680	16602	1.07	0.004	39113	82144	1.07	4.04E-08
Overweight																		
rs4735692	8	76778218	<i>HNF4G</i>	a	g	0.5839	92703	65698	1.05	6.13E-09	65323	39290	1.03	0.003	158026	104988	1.04	3.51E-10
rs7503807	17	76205706	<i>RPTOR</i>	a	c	0.5654	92855	65723	1.04	4.20E-06	64535	38813	1.03	0.0009	157390	104536	1.04	1.98E-08

*Stage 2 consists of studies with either GWAS and MetaboChip data. Not all SNPs were present on MetaboChip.

Table 2. Association results for novel genome-wide significant ($P < 5 \times 10^{-8}$) SNPs with height and obesity-related traits

SNP	Gene	Effect allele	Other allele	BMI tails		Obesity class 3		Obesity class 2		Obesity class 1		Overweight		BMI (continuous) ^a		Height tails		Height (continuous) ^a	
				OR	P	OR	P	OR	P	OR	P	OR	P	Effect	P-value	OR	P	Effect	P-value
Height tails																			
rs584438	<i>IGFBP4</i>	c	a	0.98	0.52	1.02	0.64	1.01	0.47	1.00	0.75	1.00	0.59	0.005	0.22	1.18	5.22E-12	0.025	9.43E-11
rs6662509	<i>H6PD</i>	t	c	1.00	0.95	1.11	0.07	1.01	0.83	0.99	0.34	0.99	0.35	-0.006	0.27	1.23	3.19E-10	0.031	7.76E-12
rs2362965	<i>RSRC1/SHOX2</i>	t	a	0.95	0.02	0.97	0.25	0.98	0.20	0.99	0.37	0.99	0.21	-0.007	0.05	1.12	2.14E-09	0.017	7.07E-08
rs1594829	<i>PPP2R2A</i>	c	t	1.03	0.33	1.06	0.11	1.01	0.58	1.00	0.82	1.01	0.32	0.004	0.33	1.15	3.88E-08	0.016	4.29E-05
Obesity Class 2																			
rs7989336	<i>HS6ST3</i>	a	g	1.09	0.0001	1.11	0.0006	1.10	1.06E-08	1.04	9.38E-05	1.04	2.33E-06	0.016	8.80E-06	1.00	0.89	-0.001	0.71
rs17381664	<i>ZZZ3</i>	c	t	1.08	0.0001	1.12	5.41E-05	1.09	2.85E-08	1.05	6.80E-08	1.04	2.23E-07	0.022	2.50E-11	1.06	0.005	0.010	0.004
Obesity Class 1																			
rs17024258	<i>GNAT2</i>	t	c	1.27	0.02	1.45	0.002	1.26	7.73E-05	1.25	8.66E-12	1.13	1.41E-08	0.067	4.34E-14	1.21	0.08	0.010	0.36
rs4735692	<i>HNF4G</i>	a	g	1.09	1.97E-05	1.08	0.006	1.05	0.0005	1.06	2.48E-09	1.04	3.51E-10	0.019	9.94E-10	1.02	0.50	0.006	0.10
rs13041126	<i>MRPS33P4</i>	t	c	1.08	0.001	1.05	0.16	1.06	0.0002	1.06	2.16E-08	1.04	1.43E-06	0.017	8.52E-07	1.02	0.56	0.002	0.65
rs2531995	<i>ADCY9</i>	t	c	1.06	0.01	1.09	0.006	1.06	0.001	1.07	4.04E-08	1.03	5.57E-05	0.021	6.58E-08	1.07	0.005	0.018	1.87E-06
Overweight																			
rs4735692	<i>HNF4G</i>	a	g	1.09	1.97E-05	1.08	0.006	1.05	0.0005	1.06	2.48E-09	1.04	3.51E-10	0.019	9.94E-10	1.02	0.50	0.006	0.10
rs7503807	<i>RPTOR</i>	a	c	1.08	7.07E-05	1.13	9.44E-06	1.07	2.46E-06	1.05	1.12E-07	1.04	1.98E-08	0.020	3.00E-10	0.99	0.55	-0.001	0.85

^aThe beta represents the difference in standardized effects.

Figure Legends

Figure 1. Q-Q plot of the $-\log_{10}$ p-values for the difference between the observed association for the tails of BMI and expected association based on the overall BMI distribution.

Figure 2. Variance in extreme obesity explained by common genetic variants. The phenotypic variance explained is higher when SNPs with lower degrees of significance are included in the polygenetic prediction model. The *y-axis* represents the proportion of variance explained (Nagelkerke R^2) of extreme obesity in six studies not included in the discovery meta-analysis. In panel A, the prediction model was based on the results from the stage I meta-analysis of tails of BMI. The thicker lines represent the weighted average; 95% confidence intervals are reported as double-headed arrows. In panel B, the prediction model was based on BMI from the full distribution (modified version of the previous GIANT meta-analysis by Speliotes *et al*^A). * Essen Obesity Study was not adjusted by age.

Online Methods (Word Count: 1,366)

Detailed methods descriptions are available in the Supplementary Note.

Discovery and joint meta-analyses

Study design

We conducted a two-stage study for the tails of three anthropometric traits (BMI, WHR, and height) and four clinical classes of obesity (overweight and obesity classes I, II, and III), followed by a combined analysis of the two stages. Stage 1 consisted of a meta-analysis of GWAS utilizing data from a study base (or sampling frame) of up to 168,267 adult individuals of European ancestry from 51 studies participating in the Genetic Investigation of ANthropometric Traits (GIANT) consortium (Supplementary Tables 1-5). In stage 2, 273 SNPs with P-values $< 5 \times 10^{-6}$ were followed up in up to 109,703 additional individuals of European descent, which included 67,243 individuals from 24 studies with data from the MetaboChip (a custom-designed array of $\sim 200,000$ SNPs with prior evidence of suggestive association with metabolic traits), and 42,460 individuals from 12 studies with *in silico* replication GWAS data (Supplemental Tables 1-5). This gave us a study base of up to 276,007 individuals of European descent for the joint meta-analysis of stage 1 and stage 2. For full details about the discovery and replication stages, analysis of data, and meta-analyses, see **Supplementary Note**.

Phenotype definitions

The tails of the three anthropometric traits (BMI, height, and WHR) were defined as the upper 5th percentile (cases) and lower 5th percentile (controls) of the distribution stratified by sex and disease status after controlling for the following covariates: age, age² and principal components for BMI; age and principal components for height; and age, age², BMI and principal components for WHR. For the clinical obesity classes, cases were defined as BMI ≥ 25 kg/m² for overweight, BMI ≥ 30 kg/m² for obesity class I, BMI ≥ 35 kg/m² for obesity class II, and BMI ≥ 40 kg/m² for obesity class 3. Controls were subjects with BMI < 25 kg/m². A minimum of 30 cases and 30 controls for each study-specific stratum was required.

Association analyses and meta-analyses

Each study conducted single marker association analyses assuming an additive genetic model taking genotype imputation uncertainty into account. Analyses were stratified by sex (except for studies with related individuals where analyses accounted for family structure) and disease status for studies that ascertained participants based on a relevant disease (e.g., diabetes). Before meta-analyzing data, results from each study were extensively reviewed using standardized quality control procedures to identify potential problems, such as strand issues, discrepancies between the reported standard errors and p-values, and allele frequency differences. SNPs with poor imputation quality scores or estimated minor allele count ≤ 20 (i.e. $2 \times N \times$ minor allele frequency) in each stratum (men/women or pooled for family-based studies) of each study were removed from analysis. For the discovery stage 1, each stratum- and study-specific GWAS was corrected for genomic control. Meta-analyses were performed for each phenotype in METAL⁴⁸ using the fixed effects inverse variance method based on the β estimates and standard errors from each study. The results of the discovery meta-analysis were followed by an additional genomic control correction. Similar methods were employed for the replication and joint discovery and replication analysis.

Association testing in extremely obese studies

We tested the association of all SNPs reaching $P < 5 \times 10^{-8}$ in the joint analysis of stage 1 and stage 2 results for the BMI-related traits, in five studies of extremely obese individuals (**Supplementary Tables 2-5**). For the four case-control studies (French Extreme Obesity Study, Essen Case-Control GWAS, GEO and GOYA), a fixed effects inverse variance method was used to meta-analyze the results. A fifth study (Essen Obesity Trio GWAS) that has a nuclear family structure was meta-analyzed with the four case-control studies using a weighted z-score method that takes into account the direction but not the magnitude of the association.

Systematic comparison of the genetic structure between tails and overall distribution

For these analyses, we included all GWAS studies that provided genome-wide results for both the full distribution and tails of BMI, height and WHR. First, we used the results for the full distribution to calculate, for each genotype, the expected number of individuals in the upper and lower 5% tails. We used these values to perform a logistic regression, comparing the upper and lower tails, and obtained the 'expected beta' and 'expected standard error'.

Second, we tested the differences between the 'expected betas' and the 'observed betas' obtained from the meta-analyses of the tails of the distributions. The standard error of the differences was estimated as: $\sqrt{\text{expected standard error}^2 + \text{observed standard error}^2 - 2 \cdot 0.65 \cdot (\text{expected standard error} \cdot \text{observed standard error})}$, where 0.65 is the correlation between 'expected betas' and 'observed betas' obtained from TWINGENE by bootstrapping. Finally, differences between 'expected betas' and 'observed betas' were meta-analyzed using the inverse variance method in METAL.

Polygene comparison of genetic determinants of the BMI tails and overall distribution

Within each trait (BMI, height and WHR), we aimed to compare variance explained in tails of the trait by two genetic scores (polygene scores) obtained from (1) the meta-analyses of the tails of the trait and (2) the meta-analyses of the full distribution. To make the scores comparable, we limited the polygene score construction to the studies that provided genome-wide meta-analysis results for both tails and overall distribution. After LD filtering (using $r^2 \geq 0.05$ and 1 Mb distance) and excluding SNPs present in <50% of samples, we created polygene scores, as weighted sum of risk alleles, using the method proposed by the International Schizophrenia Consortium.⁴⁵ For the BMI analysis, the association between the polygene scores and tails of BMI was investigated in four samples of extremely obese and two independent cohort studies using the same definition of tails (**Supplementary Table 13**). Only the two independent cohorts were used for the height and WHR analysis. To estimate the phenotypic variance explained, we fit logistic or linear regression models including age, sex, study-specific covariates and the polygene score as predictors, and tails of the trait or overall trait as outcomes, in separate models. The phenotypic variance explained by the polygene scores was defined as the difference in R^2 (linear regression) or Nagelkerke R^2 (logistic regression) between

these models and a basic model including only age, sex and study-specific covariates as predictors.

Secondary signals analysis

To identify potential secondary signals, we utilized the approximate conditional and joint analysis proposed by Yang et al,⁴⁶ which uses summary-level statistics and the LD structure from a reference sample to approximate conditional p-values. The meta-analysis results for each trait were analyzed separately with LD correction between SNPs estimated from 6,654 unrelated individuals from the ARIC cohort.

Haplotype-based analyses

Using data from ten of the largest studies, we tested the association between the tails of height, BMI and WHR and haplotypes across each established and novel locus separately for males and females within each study, using GENE-BPM.⁴⁹ The haplotypes were estimated from GWAS SNP data by means of an expectation-maximization algorithm and then clustered according to their allelic similarity. Within a logistic regression-modelling framework, haplotypes within the same cluster were assigned the same allelic effect, reducing the required number of parameters. Markov-chain Monte Carlo techniques were employed to sample over the space of haplotype clusters and regression model parameters. Evidence in favour of a haplotype association with the trait was assessed by summing \log_{10} Bayes factors across studies.

eQTL analyses

We examined the *cis* associations between SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) and expression of nearby genes in multiple tissues from 5 studies described previously: 1) subcutaneous adipose tissue (n=603) and whole blood (n=747) from deCode⁵⁰; 2) lymphoblastoid cell lines (n=830) from a childhood asthma study⁵¹; 3) liver (n=707), subcutaneous fat (n=870) and omental fat (n=916) tissue from a bariatric surgery study⁵²; 4) subcutaneous abdominal (N=52) and gluteal (N=62) adipose tissue and whole blood (n=65)

from MolOBB⁵³; and 5) cortical brain tissue (n=193) survey study.⁵⁴ SNPs were tested for *cis* associations with transcripts within 500 kb or 1 Mb, assuming an additive effect of the BMI allele or using an ANOVA test with study-specific p-value thresholds used to account for multiple testing. Conditional analyses were performed for all expression data, except for cortical tissue, by conditioning the trait-associated SNP on the most significant *cis*-associated SNP for that particular gene transcript and vice versa.

References

1. Maes, H.H., Neale, M.C. & Eaves, L.J. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**, 325-51 (1997).
2. Stunkard, A.J., Foch, T.T. & Hrubec, Z. A twin study of human obesity. *Jama* **256**, 51-4 (1986).
3. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin research* **6**, 399-408 (2003).
4. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
5. Okada, Y. *et al.* Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat Genet* **44**, 302-6 (2012).
6. Wen, W. *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* **44**, 307-11 (2012).
7. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949-60 (2010).
8. Lindgren, C.M. *et al.* Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* **5**, e1000508 (2009).
9. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-8 (2010).
10. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294-305 (2011).
11. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* (2012).
12. Duncan, E.L. *et al.* Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet* **7**, e1001372 (2011).
13. Edmondson, A.C. *et al.* Dense genotyping of candidate gene loci identifies variants associated with high-density lipoprotein cholesterol. *Circ Cardiovasc Genet* **4**, 145-55 (2011).
14. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* **41**, 157-9 (2009).

15. Scherag, A. *et al.* Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS Genet* **6**, e1000916 (2010).
16. Bradfield, J.P. *et al.* A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* **44**, 526-31 (2012).
17. Cohen, J.C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-72 (2004).
18. Emond, M.J. *et al.* Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet* **44**, 886-9 (2012).
19. Harismendy, O. *et al.* Population sequencing of two endocannabinoid metabolic genes identifies rare and common regulatory variants associated with extreme obesity and metabolite level. *Genome Biol* **11**, R118 (2010).
20. Romeo, S. *et al.* Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* **39**, 513-6 (2007).
21. Chan, Y. *et al.* Common variants show predicted polygenic effects on height in the tails of the distribution, except in extremely short individuals. *PLoS Genet* **7**, e1002439 (2011).
22. Cotsapas, C. *et al.* Common body mass index-associated variants confer risk of extreme obesity. *Hum Mol Genet* **18**, 3502-7 (2009).
23. Paternoster, L. *et al.* Genome-wide population-based association study of extremely overweight young adults--the GOYA study. *PLoS One* **6**, e24303 (2011).
24. Hinney, A. *et al.* Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS One* **2**, e1361 (2007).
25. Benzinou, M. *et al.* Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nat Genet* **40**, 943-5 (2008).
26. Jiao, H. *et al.* Genome wide association study identifies KCNMA1 contributing to human obesity. *BMC Med Genomics* **4**, 51 (2011).
27. den Hoed, M. *et al.* Evaluation of common genetic variants identified by GWAS for early onset and morbid obesity in population-based samples. *Int J Obes (Lond)*, E-pub ahead of print. (2012).
28. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* **41**, 25-34 (2009).
29. Putter, C. *et al.* Missing heritability in the tails of quantitative traits? A simulation study on the impact of slightly altered true genetic models. *Hum Hered* **72**, 173-81 (2011).
30. Williams, P.T. Quantile-specific penetrance of genes affecting lipoproteins, adiposity and height. *PLoS One* **7**, e28764 (2012).
31. Guey, L.T. *et al.* Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet Epidemiol*, E-pub ahead of print. (2011).
32. WHO. Obesity: preventing and managing the global epidemic. Report of a WHO Consultation. . in *WHO Technical Report Series 894* (World Health Organization, Geneva, 2000).

33. Kumanyika, S.K. *et al.* Population-based prevention of obesity: the need for comprehensive promotion of healthful eating, physical activity, and energy balance: a scientific statement from American Heart Association Council on Epidemiology and Prevention, Interdisciplinary Committee for Prevention (formerly the expert panel on population and prevention science). *Circulation* **118**, 428-64 (2008).
34. Sarbassov, D.D. & Sabatini, D.M. Redox regulation of the nutrient-sensitive raptor-mTOR pathway and complex. *J Biol Chem* **280**, 39505-9 (2005).
35. Daigo, K. *et al.* Proteomic analysis of native hepatocyte nuclear factor-4alpha (HNF4alpha) isoforms, phosphorylation status, and interactive cofactors. *J Biol Chem* **286**, 674-86 (2011).
36. Nakajima, H. *et al.* Hepatocyte nuclear factor-4 alpha gene mutations in Japanese non-insulin dependent diabetes mellitus (NIDDM) patients. *Res Commun Mol Pathol Pharmacol* **94**, 327-30 (1996).
37. Cho, Y.S. *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* **44**, 67-72 (2012).
38. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
39. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* **42**, 142-8 (2010).
40. Shi, J. & Kandror, K.V. Sortilin is essential and sufficient for the formation of Glut4 storage vesicles in 3T3-L1 adipocytes. *Dev Cell* **9**, 99-108 (2005).
41. Kaddai, V. *et al.* Involvement of TNF-alpha in abnormal adipocyte and muscle sortilin expression in obese mice and humans. *Diabetologia* **52**, 932-40 (2009).
42. Zhang, M. *et al.* Paracrine overexpression of IGFBP-4 in osteoblasts of transgenic mice decreases bone turnover and causes global growth retardation. *Journal of bone and mineral research* **18**, 836-43 (2003).
43. Liao, Y.C., Chen, N.T., Shih, Y.P., Dong, Y. & Lo, S.H. Up-regulation of C-terminal tensin-like molecule promotes the tumorigenicity of colon cancer through beta-catenin. *Cancer Res* **69**, 4563-6 (2009).
44. Milat, F. & Ng, K.W. Is Wnt signalling the final common pathway leading to bone formation? *Mol Cell Endocrinol* **310**, 52-62 (2009).
45. Purcell, S.M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).
46. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75 (2012).
47. Kilpelainen, T.O., Bingham, S.A., Khaw, K.T., Wareham, N.J. & Loos, R.J. Association of variants in the PCSK1 gene with obesity in the EPIC-Norfolk study. *Hum Mol Genet* **18**, 3496-501 (2009).
48. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
49. Morris, A.P. Direct analysis of unphased SNP genotype data in population-based association studies via Bayesian partition modelling of haplotypes. *Genet Epidemiol* **29**, 91-107 (2005).

50. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423-8 (2008).
51. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat Genet* **39**, 1202-7 (2007).
52. Zhong, H., Yang, X., Kaplan, L.M., Molony, C. & Schadt, E.E. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet* **86**, 581-91 (2010).
53. Min, J.L. *et al.* Coexpression Network Analysis in Abdominal and Gluteal Adipose Tissue Reveals Regulatory Genetic Loci for Metabolic Syndrome and Related Phenotypes. *PLoS Genet* **8**, e1002505 (2012).
54. Myers, A.J. *et al.* A survey of genetic human cortical gene expression. *Nat Genet* **39**, 1494-9 (2007).

Affiliations

1. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892, USA
2. Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala, Sweden
3. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden
4. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
5. Estonian Genome Center, University of Tartu, Tartu 50410, Estonia
6. Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK
7. Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA
8. Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA
9. Center for Observational Research, Amgen, Thousand Oaks, CA, 91320
10. Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA
11. MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK
12. Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia
13. Molecular Biology Department, Istituto Auxologico Italiano, Milano, Italy
14. Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA
15. Divisions of Genetics and Endocrinology and Center for Basic and Translational Obesity Research, Children's Hospital, Boston, Massachusetts 02115, USA

16. Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA
17. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA
18. Department of Internal Medicine (Cardiovascular), University of Michigan, Ann Arbor, MI 48109, USA
19. Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA
20. Department of Computational Medicine and Bioinformatics, University of Michigan, MI 48109, USA
21. Public Health and Gender Studies, Institute of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany
22. Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, EX1 2LU, UK
23. Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA
24. Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA
25. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia 30322, USA
26. The Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia
27. Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA
28. Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA
29. Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA
30. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114
31. deCODE Genetics, 101 Reykjavik, Iceland
32. University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland 4102, Australia
33. Queensland Institute of Medical Research, Brisbane 4029, Australia
34. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany
35. Department of Epidemiology, Erasmus MC, Rotterdam, 3015GE, The Netherlands
36. Genetic Epidemiology and Biostatistics Platform, Ontario Institute for Cancer Research. Toronto, Canada, M5G 1L7
37. Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Toronto, Canada, M5G 1X5

38. Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Crawley, Western Australia 6009, Australia
39. Department of Internal Medicine, VU University Medical Centre, Amsterdam, The Netherlands
40. National Institute for Health and Welfare, Department of Chronic Disease Prevention, Unit of Public Health Genomics, 00014, Helsinki, Finland
41. Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK
42. Department of Biological Psychology, VU University Amsterdam, 1081 BT Amsterdam, The Netherlands
43. MRC Human Genetics Unit, MRC Institute for Genetics and Molecular Medicine, Western General Hospital, Edinburgh, EH4 2XU, UK
44. Institute of Medical Biometry and Epidemiology, University of Marburg, 35037 Marburg, Germany
45. Department of Immunology, Genetics and Pathology, Uppsala University, Sweden
46. Uppsala Clinical Research Center, Uppsala university hospital, Sweden
47. Genome Centre, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK
48. Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary, University of London, London, UK
49. LURIC Study nonprofit LLC, Freiburg, Germany
50. Mannheim Institute of Public Health, Social and Preventive Medicine, Medical Faculty of Mannheim, University of Heidelberg, Mannheim, Germany
51. Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, 23562 Lübeck, Germany
52. Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland
53. Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
54. Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria
55. University Lille Nord de France, 59000 Lille, France
56. CNRS UMR8199-IBL-Institut Pasteur de Lille, F-59000 Lille, France
57. Cardiovascular Health Research Unit, University of Washington, Seattle, Washington 98101, USA
58. Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK
59. School of Social and Community Medicine, University of Bristol, UK
60. Department of Internal Medicine, Erasmus MC, Rotterdam, 3015GE, The Netherlands

61. Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA)
62. Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany
63. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology and Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
64. Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA
65. Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands
66. MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol, BS8 2BN, UK
67. Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital of Essen, University of Duisburg-Essen, Essen, Germany
68. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, 00014, Helsinki, Finland
69. Universität zu Lübeck, Medizinische Klinik II, 23538 Lübeck, Germany
70. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA
71. Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Skåne University Hospital Malmö, Lund University, Malmö, Sweden
72. Department of Public Health & Clinical Medicine, Umeå University, Umeå, Sweden
73. Department of Odontology, Umeå University, Sweden
74. Icelandic Heart Association, Kopavogur, Iceland
75. Department of Medicine, University of Iceland, Reykjavik, Iceland
76. Atherosclerosis Research Unit, Department of Medicine, Solna, Karolinska Institutet, Karolinska University Hospital, 171 76 Stockholm, Sweden
77. Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany
78. Department of Vascular Medicine, Academic Medical Center, Amsterdam, The Netherlands
79. Heart Failure Research Centre, Department of Clinical and Experimental Cardiology, Academic Medical Center, Amsterdam, the Netherlands
80. Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK
81. Department of Endocrinology, University Medical Center Groningen, University of Groningen, P.O. Box 30001, 9700 RB Groningen, The Netherlands
82. LifeLines Cohort Study, University Medical Center Groningen, University of Groningen, The Netherlands

83. Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 413 45 Gothenburg, Sweden
84. Hudson Alpha Institute for Biotechnology, Huntsville, Alabama 35806, USA
85. Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, The Netherlands
86. Institute of Biomedicine/Physiology, University of Eastern Finland, Kuopio Campus, Finland
87. NIHR Cambridge Biomedical Research Centre, Cambridge, UK
88. Division of Epidemiology, Multidisciplinary Cardiovascular Research Centre (MCRC), Leeds Institute of Genetics, Health and Therapeutics (LIGHT), University of Leeds, Leeds LS2 9JT, UK
89. PathWest Laboratory of Western Australia, Department of Molecular Genetics, J Block, QEII Medical Centre, Nedlands, Western Australia 6009, Australia
90. Department of Surgery and Pathology, University of Western Australia, Nedlands, Australia, 6009
91. Genome Technology Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA
92. Dipartimento di Medicina Sperimentale. Università degli Studi Milano-Bicocca, Monza, Italy
93. Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland
94. Harvard Medical School, Boston, Massachusetts 02115, USA
95. British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, G12 8TA, UK
96. University of Dundee, Ninewells Hospital & Medical School, Dundee, DD1 9SY, UK
97. National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK
98. Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
99. Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands
100. Istituto di Ricerca Genetica e Biomedica del CNR, Monserrato, 09042, Cagliari, Italy
101. Department of Dietetics-Nutrition, Harokopio University, 70 El. Venizelou Str, Athens, Greece
102. Epidemiology and Preventive Medicine Research Center, Department of Clinical and Experimental Medicine, University of Insubria, Varese, Italy
103. Department of Cardiology, Toulouse University School of Medicine, Rangueil Hospital, Toulouse, France

104. Department of Genetics, University Medical Center Groningen, University of Groningen, The Netherlands
105. University of Milan, Department of Health Sciences, Ospedale San Paolo, 20139 Milano, Italy
106. University of Chicago, Chicago, Illinois 60637, USA
107. Northshore University HealthSystem, Evanston, Illinois 60201, USA
108. Research Unit for Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
109. Division of Cardiovascular and Neuronal Remodelling, Multidisciplinary Cardiovascular Research Centre, Leeds Institute of Genetics, Health and Therapeutics, University of Leeds, UK
110. Department of Clinical Sciences/Obstetrics and Gynecology, University of Oulu, 90014 Oulu, Finland
111. Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118, USA
112. Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63108, USA
113. Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, 45147 Essen, Germany
114. Centre For Paediatric Epidemiology and Biostatistics/MRC Centre of Epidemiology for Child Health, University College of London Institute of Child Health, London, UK
115. Division of Research, Kaiser Permanente Northern California, Oakland, California 94612, USA
116. Core Genotyping Facility, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702, USA
117. Department of Physiology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, 405 30 Gothenburg, Sweden
118. National Institute for Health and Welfare, Department of Chronic Disease Prevention, Population Studies Unit, 20720 Turku, Finland
119. Department of Clinical Physiology, University of Tampere and Tampere University Hospital, 33520 Tampere, Finland;
120. Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
121. Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
122. Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA
123. UKCRC Centre of Excellence for Public Health (NI) Queens University, Belfast

124. Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge CB2 2SR, UK
125. Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, UK
126. Department of Internal Medicine II – Cardiology, University of Ulm Medical Center, Ulm, Germany
127. National Institute for Health and Welfare, Department of Chronic Disease Prevention, Chronic Disease Epidemiology and Prevention Unit, 00271, Helsinki, Finland
128. Department of Medicine, University of Eastern Finland, Kuopio Campus and Kuopio University Hospital, 70210 Kuopio, Finland
129. Finnish Institute of Occupational Health, 90220 Oulu, Finland
130. Kuopio Research Institute of Exercise Medicine, Kuopio, Finland
131. Laboratory of Epidemiology, Demography, Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA
132. Department of Medical Sciences, Uppsala University, Akademiska sjukhuset, 751 85 Uppsala, Sweden
133. National Institute for Health and Welfare, Diabetes Prevention Unit, 00271 Helsinki, Finland
134. Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore
135. Department of Internal Medicine, Istituto Auxologico Italiano, Verbania, Italy
136. Transplantation Laboratory, Haartman Institute, University of Helsinki, 00014, Helsinki, Finland
137. Università Vita-Salute San Raffaele, Chair of Nephrology San Raffaele Scientific Institute, OU Nephrology and Dialysis, 20132 Milan, Italy
138. Synlab Academy, Mannheim, Germany
139. Department of Cardiology, University Medical Center Groningen, University of Groningen, The Netherlands
140. Departments of Biostatistics, University of Washington, Seattle, Washington 98195, USA
141. Genetics Division, GlaxoSmithKline, King of Prussia, Pennsylvania 19406, USA
142. Institute of Human Genetics, University of Bonn, Bonn, Germany
143. Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
144. School of Population Health, The University of Western Australia, Nedlands WA 6009, Australia
145. Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Australia, 6009
146. Busselton Population Medical Research Foundation Inc., Sir Charles Gairdner Hospital, Nedlands, Western Australia 6009, Australia

147. Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen
148. MRC Harwell, Harwell, UK
149. Department of Statistics, University of Oxford, Oxford OX1 3TG, UK
150. Department of Public Health, Section of Epidemiology, Aarhus University, Denmark
151. MRC Unit for Lifelong Health & Ageing, London, UK
152. Department of Clinical Genetics, Erasmus MC, Rotterdam, 3015GE, The Netherlands
153. Centre for Medical Systems Biology & Netherlands Consortium on Healthy Aging, Leiden, the Netherlands
154. Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School. Dundee, DD1 9SY
155. Illumina Inc. Cambridge
156. Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
157. Munich Heart Alliance, Munich, Germany
158. Faculty of Medicine, University of Split, Croatia
159. National Institute for Health and Welfare, 90101 Oulu, Finland
160. Center for Biomedicine, European Academy Bozen/Bolzano (EURAC), Bolzano/Bozen, 39100, Italy - Affiliated Institute of the University of Lübeck, Lübeck, Germany.
161. Department of Neurology, General Central Hospital, Bolzano, Italy
162. Department of Neurology, University of Lübeck, Lübeck, Germany.
163. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, OX3 7LJ, UK
164. Department of Haematology, University of Cambridge, Cambridge CB2 0PT, UK
165. NHS Blood and Transplant, Cambridge Centre, Cambridge, CB2 0PT, UK
166. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, 20520 Turku, Finland
167. The Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, 20520 Turku, Finland
168. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK
169. Finnish Diabetes Association, Kirjoniementie 15, 33680, Tampere, Finland
170. Pirkanmaa Hospital District, Tampere, Finland
171. South Karelia Central Hospital, 53130 Lappeenranta, Finland
172. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
173. Institute for Clinical Molecular Biology, Christian-Albrechts University, Kiel, Germany
174. Deutsches Zentrum für Herz-Kreislaufforschung e. V. (DZHK), Universität zu Lübeck, 23538 Lübeck, Germany
175. Azienda ospedaliera di Desio e Vimercate, Milano, Italy

176. Division of Cardiology, Cardiovascular Laboratory, Helsinki University Central Hospital, 00029 Helsinki, Finland
177. University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland
178. Klinik und Poliklinik für Innere Medizin II, Universitätsklinikum Regensburg, 93053 Regensburg, Germany
179. Department of Medicine, University of Leipzig, 04103 Leipzig, Germany
180. University of Leipzig, IFB Adiposity Diseases, Leipzig, Germany
181. INSERM UMR_S 937, ICAN Institute, Pierre et Marie Curie Medical School, Paris 75013, France
182. Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano, Centro Cardiologico Monzino, IRCCS, Milan, Italy
183. Department of Human Genetics, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands
184. Department of Medicine, University of Turku and Turku University Hospital, 20520 Turku, Finland
185. Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV) University Hospital, 1011 Lausanne, Switzerland
186. Cardiology Group, Frankfurt-Sachsenhausen, Germany
187. Institut Pasteur de Lille, INSERM U744, Université Lille Nord de France, F-59000 Lille, France
188. Division of Endocrinology and Diabetes, Department of Medicine, University Hospital, Ulm, Germany
189. Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas 77030, USA
190. Fondazione Filarete, Milano, Italy
191. Department of General Practice and Primary health Care, University of Helsinki, Helsinki, Finland
192. National Institute for Health and Welfare, 00271 Helsinki, Finland
193. Helsinki University Central Hospital, Unit of General Practice, 00280 Helsinki, Finland
194. Department of Nutrition, Harvard School of Public Health, Boston, MA
195. Department of Genomics of Common Disease, School of Public Health, Imperial College London, W12 0NN, London, UK
196. HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway
197. Hannover Unified Biobank, Hannover Medical School, 30625 Hannover, Germany
198. Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, W2 1PG, UK
199. Institute of Health Sciences, University of Oulu, 90014 Oulu, Finland

200. Biocenter Oulu, University of Oulu, 90014 Oulu, Finland
201. Unit of General Practice, Oulu University Hospital, Oulu, Finland
202. Department of Urology, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands
203. Comprehensive Cancer Center East, 6501 BG Nijmegen, The Netherlands
204. Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere and Tampere University Hospital, 33520 Tampere, Finland
205. Stanford University School of Medicine, Stanford, California 93405, USA
206. Department of Clinical Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway
207. Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway
208. Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, The Netherlands
209. Department of Psychiatry, University Medical Centre Groningen, 9713 GZ Groningen, The Netherlands
210. Departments of Epidemiology, Medicine and Health Services, University of Washington, Seattle, Washington 98195, USA
211. Group Health Research Institute, Group Health, Seattle, Washington 98101, USA
212. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland
213. Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, LE3 9QP, UK
214. Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, LE3 9QP, UK
215. Institute of Preventive Medicine, Bispebjerg University Hospital, Copenhagen, and Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark
216. Faculty of Medicine, University of Iceland, 101 Reykjavík, Iceland
217. Red RECAVA Grupo RD06/0014/0015, Hospital Universitario La Paz, 28046 Madrid, Spain
218. Centre for Vascular Prevention, Danube-University Krems, 3500 Krems, Austria
219. South Ostrobothnia Central Hospital, 60220 Seinajoki, Finland
220. Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Finland
221. Research Unit, Kuopio University Hospital, Kuopio, Finland
222. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany

223. Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
224. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
225. Klinikum Grosshadern, Munich, Germany
226. Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, USA
227. University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke's Hospital, CB2 0QQ, Cambridge, UK
228. Division of Intramural Research, National Heart, Lung and Blood Institute, Framingham Heart Study, Framingham, Massachusetts 01702, USA
229. Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, 20502 Malmö, Sweden
230. Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA
231. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA
232. Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Old Road Headington, Oxford, OX3 7LJ, UK
233. Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA
234. Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1498, New York, NY 10029-6574 USA
235. Institute of Genomics and Multiscale Biology, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1498, New York, NY 10029-6574 USA
236. Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland 21224, USA
237. Division of Population Health Sciences and Education, St George's, University of London, London, SW17 0RE, UK
238. Center of Medical Systems Biology, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands
239. Department of Clinical Epidemiology and Biostatistics, McMasterUniversity, Hamilton, Ontario L8S 4L8, Canada
240. Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA
241. Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA
242. The Charles Bronfman Institute of Personalized Medicine, Mount Sinai School of Medicine, New York, NY 10029, USA

243. The Mindich Child Health and Development Institute, Mount Sinai School of Medicine, New York, NY 10029, USA
244. Department of Preventive Medicine, Mount Sinai School of Medicine, New York, NY 10029, USA