

EPIDEMIOLOGY

Tissue-specific genetic variation suggests distinct molecular pathways between body shape phenotypes and colorectal cancer

Laia Peruchet-Noray^{1,2}, Anja M. Sedlmeier^{3,4,5}, Niki Dimou¹, Hansjörg Baurecht⁵, Béatrice Fervers⁶, Emma Fontvieille¹, Julian Konzok⁵, Kostas K. Tsilidis^{7,8}, Sofia Christakoudi^{7,9}, Anna Jansana¹, Reynalda Cordova^{1,10}, Patricia Bohmann⁵, Michael J. Stein⁵, Andrea Weber⁵, Stéphane Bézieau¹¹, Hermann Brenner^{12,13,14}, Andrew T. Chan¹⁵, Iona Cheng¹⁶, Jane C. Figueiredo¹⁷, Koldo Garcia-Etxebarria^{18,19}, Victor Moreno^{2,20,21,22}, Christina C. Newton²³, Stephanie L. Schmit^{24,25}, Mingyang Song^{15,26}, Cornelia M. Ulrich²⁷, Pietro Ferrari¹, Vivian Viallon¹, Robert Carreras-Torres²⁸†*, Marc J. Gunter^{1,7}†*, Heinz Freisling¹†*

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

It remains unknown whether adiposity subtypes are differentially associated with colorectal cancer (CRC). To move beyond single-trait anthropometric indicators, we derived four multi-trait body shape phenotypes reflecting adiposity subtypes from principal components analysis on body mass index, height, weight, waist-to-hip ratio, and waist and hip circumference. A generally obese (PC1) and a tall, centrally obese (PC3) body shape were both positively associated with CRC risk in observational analyses in 329,828 UK Biobank participants (3728 cases). In genome-wide association studies in 460,198 UK Biobank participants, we identified 3414 genetic variants across four body shapes and Mendelian randomization analyses confirmed positive associations of PC1 and PC3 with CRC risk (52,775 cases/45,940 controls from GECCO/CORECT/CCFR). Brain tissue-specific genetic instruments, mapped to PC1 through enrichment analysis, were responsible for the relationship between PC1 and CRC, while the relationship between PC3 and CRC was predominantly driven by adipose tissue-specific genetic instruments. This study suggests distinct putative causal pathways between adiposity subtypes and CRC.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer death worldwide with more than 1.9 million incident cases and nearly 1 million deaths in 2020 (1, 2). There is convincing evidence that individuals with overweight or obesity [body mass index (BMI) ≥ 25 kg/m²] have a higher risk of CRC (3). Waist or hip circumferences (WC or HC) and waist-to-hip ratio (WHR) represent surrogate markers of body fat distribution and have shown associations with CRC risk similar to BMI, without providing additional insight into CRC etiology or risk discrimination (4). Jointly interrogating these anthropometric traits may provide information on the role of body shapes in cancer development

and advance knowledge on how adiposity subtypes are differentially associated with CRC risk.

A promising way to define body shape was proposed by Ried *et al.* in 2016 (5). This approach was derived from data on 170,000 individuals of European ancestry and was based on a principal components analysis (PCA) on BMI, weight, height, WC, HC, and WHR. We reported previously that body shape phenotypes were associated with a higher risk of 17 different cancers in an observational analysis in the European Prospective Investigation into Cancer and Nutrition (EPIC) (6). Two distinct and orthogonal body shapes characterizing a generally obese and a tall, centrally obese body shape, respectively, were both strongly positively associated with CRC risk

¹International Agency for Research on Cancer (IARC/WHO), Nutrition and Metabolism Branch, 69366 Lyon CEDEX 07, France. ²Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain. ³Center for Translational Oncology, University Hospital Regensburg, Regensburg, Germany. ⁴Bavarian Cancer Research Center (BZKF), Regensburg, Germany. ⁵Department of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany. ⁶Département Prévention Cancer Environnement, Centre Léon Bérard, Lyon, France. ⁷Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK. ⁸Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece. ⁹Department of Inflammation Biology, School of Immunology & Microbial Sciences, King's College London, London, UK. ¹⁰Department of Nutritional Sciences, University of Vienna, Vienna, Austria. ¹¹Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France. ¹²Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹³Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. ¹⁴German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁵Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ¹⁶Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA. ¹⁷Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ¹⁸Biodonostia, Gastrointestinal Genetics Group, San Sebastián, Spain. ¹⁹Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain. ²⁰Unit of Biomarkers and Susceptibility (UBS), Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO), L'Hospitalet del Llobregat, 08908 Barcelona, Spain. ²¹ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 08908 Barcelona, Spain. ²²Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain. ²³Department of Population Science, American Cancer Society, Atlanta, GA, USA. ²⁴Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH, USA. ²⁵Population and Cancer Prevention Program, Case Comprehensive Cancer Center, Cleveland, OH, USA. ²⁶Departments of Epidemiology and Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA. ²⁷Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA. ²⁸Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), Salt, Girona, Spain.

†These authors contributed equally to this work.

*Corresponding author. Email: freisling@iarc.who.fr (H.F.); m.gunter@imperial.ac.uk (M.J.G.); rcarreras@idibgi.org (R.C.-T.)

in the observational analysis in EPIC (6). These initial findings require replication in a different study population and assessment of their potential causal basis. Moreover, colorectal carcinogenesis is characterized by a variety of genetic and molecular changes (7), and body shapes may operate through such distinct, but currently unknown, molecular pathways.

The genetic analysis of the body shapes identified genetic variants not previously associated with anthropometric traits in isolation (5). However, several genetic variants that were identified in the latest single-trait genome-wide association studies (GWASs) were not found in the study by Ried *et al.* (5). Thus, an updated GWAS in a larger sample of study participants could find additional genetic variants and provide more accurate estimates of the associations with body shapes.

Mendelian randomization (MR) studies use genetic variants as instrumental variables for a given exposure to provide evidence of potential causality in exposure-outcome associations (8). The key sources of bias in MR studies are unrelated to those of observational analyses, which means that both approaches complement each other and could thus provide more robust evidence of causality via triangulation (9). In addition, genetic variants can be grouped through enrichment analysis based on the tissue in which their mapped genes show different expression levels as compared to other tissues. Enrichment (i.e., differential expression) of BMI- and WHR-associated genes has been observed in tissues of the central nervous system and of the digestive and urogenital systems, respectively (10, 11).

Trait-associated genes that show differential expression across tissues may reflect different tissue contribution to a phenotype with distinct effects on disease risk (12). Such an approach has the potential to identify more specific molecular targets to intervene on the obesity-CRC relationship.

This analysis aimed to triangulate the association between four distinct body shapes and the risk of CRC, by subsite (colon, distal colon, proximal colon, and rectum) and by sex, and to unravel potential mechanisms by examining the role of tissue-specific gene expression of the body shapes in CRC development. We performed (i) a prospective observational analysis in 329,828 UK Biobank participants (3728 incident CRC cases), (ii) a genome-wide association of 460,198 UK Biobank participants and a tissue expression enrichment analysis in the Genotype-Tissue Expression project (GTEx) V8 dataset to identify functional genetic instruments of body shapes, and (iii) an MR analysis using data from large CRC genetic consortia including 98,715 participants of European ancestry (52,775 cases and 45,940 controls) (Fig. 1).

RESULTS

Body shape phenotypes

The first four PCs as derived in UK Biobank participants captured more than 99% of the variation in BMI, height, weight, WHR, WC, and HC (table S1), defining four distinct body shapes with high consistency to previous studies (5, 6), ancestries (fig. S1), and by sex

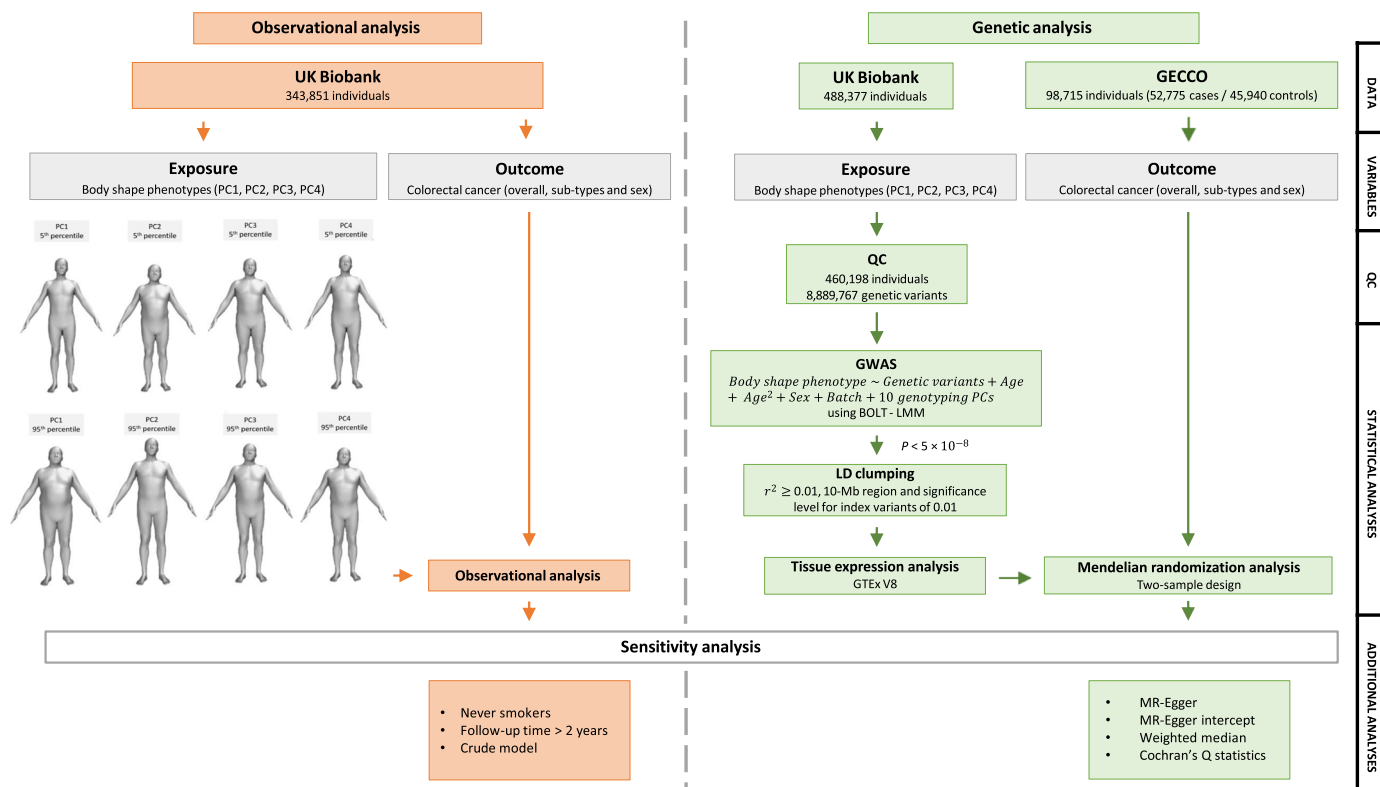


Fig. 1. Flowchart summarizing study methods. Body shape phenotypes have been derived by a PCA on six anthropometric traits (BMI, weight, height, WHR, WC, and HC). PC1 showed high and same sign loadings for all traits except height. PC2 showed high but opposite loadings for height and WHR. PC3 was characterized by high and same direction loadings for height and WHR. PC4 showed high loadings for weight and BMI and low loadings for HC and WC. QC, quality control.

(table S2). The first PC (PC1) accounted for 66.2% of total variation of the six traits describing a generally obese body shape. PC2 explained 19.4% of total variation defining tall individuals with low WHR. PC3 explained 12.3% of total variation characterizing a tall, centrally obese body shape. PC4 explained 2.0% of total variation defining an athletic body shape. Figure 1 shows these four body shapes graphically.

Observational analysis of body shape phenotypes and CRC risk

We investigated whether the four body shapes were associated with the risk of CRC in an observational analysis in UK Biobank. Baseline characteristics of the study population by sex-specific quintiles of PC1 are presented in table S3. Men and women in the lowest quintile (versus the highest quintile) of PC1 tended to have lower weight, height, BMI, WC, HC, and WHR; used less nonsteroidal anti-inflammatory drugs (NSAIDs); were more physically active; were less sedentary; tended to be never smokers; and had higher levels of education and a healthier diet. During a median follow-up time of 10.9 years (interquartile range: 10.1 to 11.6 years; 3,467,670 person-years), a total of 3728 incident CRC cases were recorded. The multivariable-adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) per 1-SD increment of PC1 was 1.13 (95% CI: 1.09 to 1.17), with similar results among men (HR: 1.17; 95% CI: 1.11 to 1.23) and women (HR: 1.09; 95% CI: 1.03 to 1.15) (P value for interaction by sex = 0.06). Positive associations were found for colon cancer (HR: 1.16; 95% CI: 1.11 to 1.21), regardless of its subsites (distal colon: HR: 1.15; 95% CI: 1.07 to 1.23; proximal colon: HR: 1.17; 95% CI: 1.11 to 1.24; P value for subsites heterogeneity = 0.10), and rectal cancer (HR: 1.07; 95% CI: 1.01 to 1.14). PC3 was also associated with an increased CRC risk with an HR of 1.09 per 1-SD increment (95% CI: 1.05 to 1.13), which was similar between men (HR: 1.09; 95% CI: 1.04 to 1.15) and women (HR: 1.08; 95% CI: 1.03 to 1.13), and for cancers of the colon (HR: 1.10; 95% CI: 1.06 to 1.14) and rectum (HR: 1.07; 95% CI: 1.01 to 1.14). PC2 and PC4 were not associated with CRC risk (Fig. 2 and table S4). Sensitivity analyses excluding current and former smokers, the two initial years of follow-up, and a crude model that included all four PCs supported the risk estimates of the main model for all four body shapes (table S4).

Genome-wide association study

We performed a GWAS in 460,198 UK Biobank participants for each of the four body shapes. We identified 3414 independent genetic variants that reached genome-wide significance ($P < 5 \times 10^{-8}$) after linkage disequilibrium (LD) clumping (see Materials and Methods). Of these, 570 variants were not previously associated with any of the six single anthropometric traits that built the body shapes; 59 (1.7%) genetic variants were shared between two or more body shapes (fig. S2). Manhattan plots, Q-Q plots, and the summary statistic results can be found in Fig. 3, fig. S3, and data S1, respectively.

For PC1, we identified 678 genetic variants ($P < 5 \times 10^{-8}$), of which 21 were not previously linked to the six anthropometric traits. Among these genetic variants, rs10803149, in the *AKT3* gene (chromosome 1), was strongly associated with PC1 (β A versus T = 0.039; $P = 2.00 \times 10^{-10}$). This gene is of interest because it encodes a protein of the serine/threonine protein kinase family, which has a key role in regulating cell survival, insulin signaling, angiogenesis, and tumor formation (13, 14). Another lead genetic variant strongly associated

with PC1 (rs1421085, β T versus C = -0.140 ; $P = 1.90 \times 10^{-241}$) in the *FTO* gene (chromosome 16) has been previously linked to BMI, weight, WC and HC, and WHR (15–19). The *FTO* gene acts as a regulator of fat mass, adipogenesis, and energy homeostasis (13, 14).

For PC2, of a total of 1802 genetic variants ($P < 5 \times 10^{-8}$), 224 were not previously linked to the six anthropometric traits. Among these 224, the most strongly associated was rs9270074 (β T versus C = 0.040; $P = 2.4 \times 10^{-41}$), an intronic variant in the *HLA-DRB1* gene (chromosome 6). This gene is likely providing antitumor immunity in several solid and hematological malignancies (13, 14). Of all genetic variants associated with PC2, rs6440003 and rs143384 were most strongly associated (β G versus A = -0.086 ; $P < 1.4 \times 10^{-217}$ and β A versus G = -0.092 ; $P < 1.4 \times 10^{-217}$, respectively). Rs6440003 is in the gene *ZBTB38* (chromosome 3), which plays an important role in regulating DNA replication (13, 14), while rs143384 is in the *GDF5* gene (chromosome 20), which encodes a growth factor involved in bone and cartilage formation (13, 14).

For PC3, we identified 741 genetic variants ($P < 5 \times 10^{-8}$), of which 220 were not previously linked to the six anthropometric traits. Among these 220 genetic variants, rs114496946, located in the gene *RAF1* (β T versus C = 0.077; $P = 1.50 \times 10^{-24}$), which encodes a serine/threonine protein kinase that regulates cell fate decisions including proliferation, differentiation, apoptosis, survival, and oncogenic transformation (13, 14). The genetic variant rs72959041 (β G versus A = -0.149 ; $P = 1.20 \times 10^{-277}$) is located in the gene *RSPO3* (chromosome 6), which is involved in the regulation of angiogenesis (13, 14). This variant was the strongest associated genetic variant with waist-to-hip index and WHR adjusted for BMI in a recent GWAS on allometric indices (20).

For PC4, we identified 193 genetic variants ($P < 5 \times 10^{-8}$), of which 105 were not previously known for these six anthropometric traits. Among these, rs12952730 (β G versus A = -0.005 ; $P = 1.50 \times 10^{-12}$) in the *RPS6KB1* gene (chromosome 17) encodes a mammalian target of rapamycin (mTOR) responding protein signaling to promote protein synthesis, cell growth, and cell proliferation. This gene has been associated with human cancer (13, 14). Another genetic variant strongly associated with PC4 was rs9907540 (β A versus G = -0.009 ; $P = 9.90 \times 10^{-36}$), located in the *VPS53* gene (chromosome 17), which is involved in cellular retrograde transport from early and late endosomes to the trans-Golgi network and in endocytic recycling (13, 14).

Overall, 192 of 207 (92.75%) genetic variants associated with the body shapes reported by Ried *et al.* were also identified in our GWAS (27 of 31 for PC1, 115 of 124 for PC2, 44 of 45 for PC3, and 6 of 7 for PC4). The genetic effects (measured by GWAS betas) of genetic variants associated with PC1 were larger than for genetic variants associated with the other PCs (Fig. 4A). This was reflected in the proportion of variability in the body shapes explained by these independent genetic variants, which were 30.9% for PC1, 34.8% for PC2, 7.8% for PC3, and 0.2% for PC4.

Tissue expression enrichment analysis

A tissue expression enrichment analysis of the body shape GWAS results was carried out using the Multi-marker Analysis of GenoMic Annotation (MAGMA) gene set analysis. This test uses the full distribution of GWAS P values to test for gene expression enrichment in 54 specific tissues across 30 general tissues of the GTEx V8 dataset (see Materials and Methods). The results are shown in Fig. 4B and fig. S4. Briefly, PC1 was enriched for gene expression in the

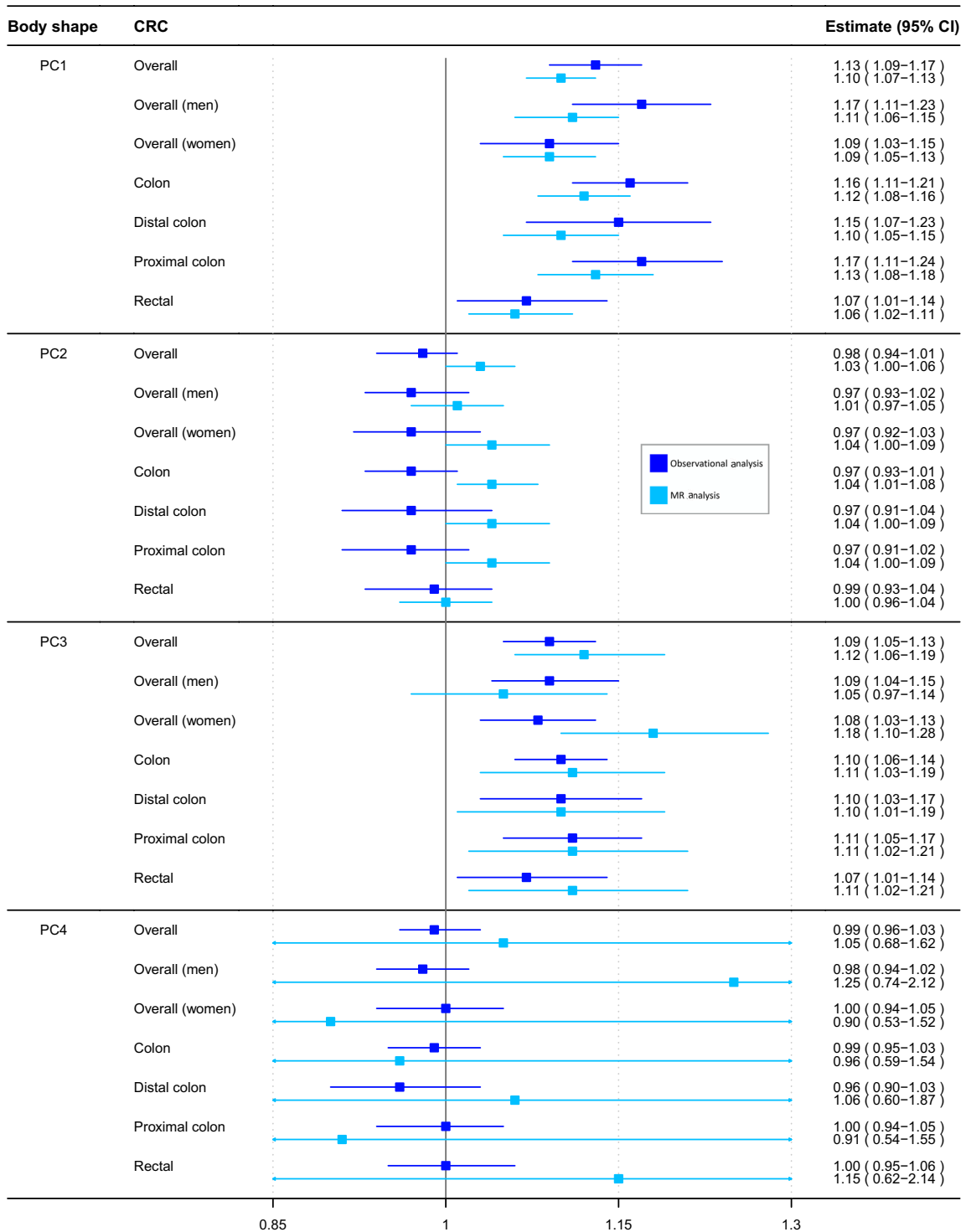


Fig. 2. Observational (dark blue) and MR (light blue) estimates of associations between body shape phenotype PC1, PC2, PC3, and PC4, per 1 standard deviation increments, and the risk of CRC (overall, among men, among women, and CRC subsites). CI, confidence interval; CRC, colorectal cancer; PC, principal component. For observational and MR analyses, cancer cases were 3728 and 52,775 for overall CRC, 2239 and 28,207 among men, 1489 and 24,568 among women, 2443 and 27,817 colon, 1285 and 13,713 rectal, 966 and 14,016 distal colon, and 1370 and 12,360 proximal colon subsites, respectively.

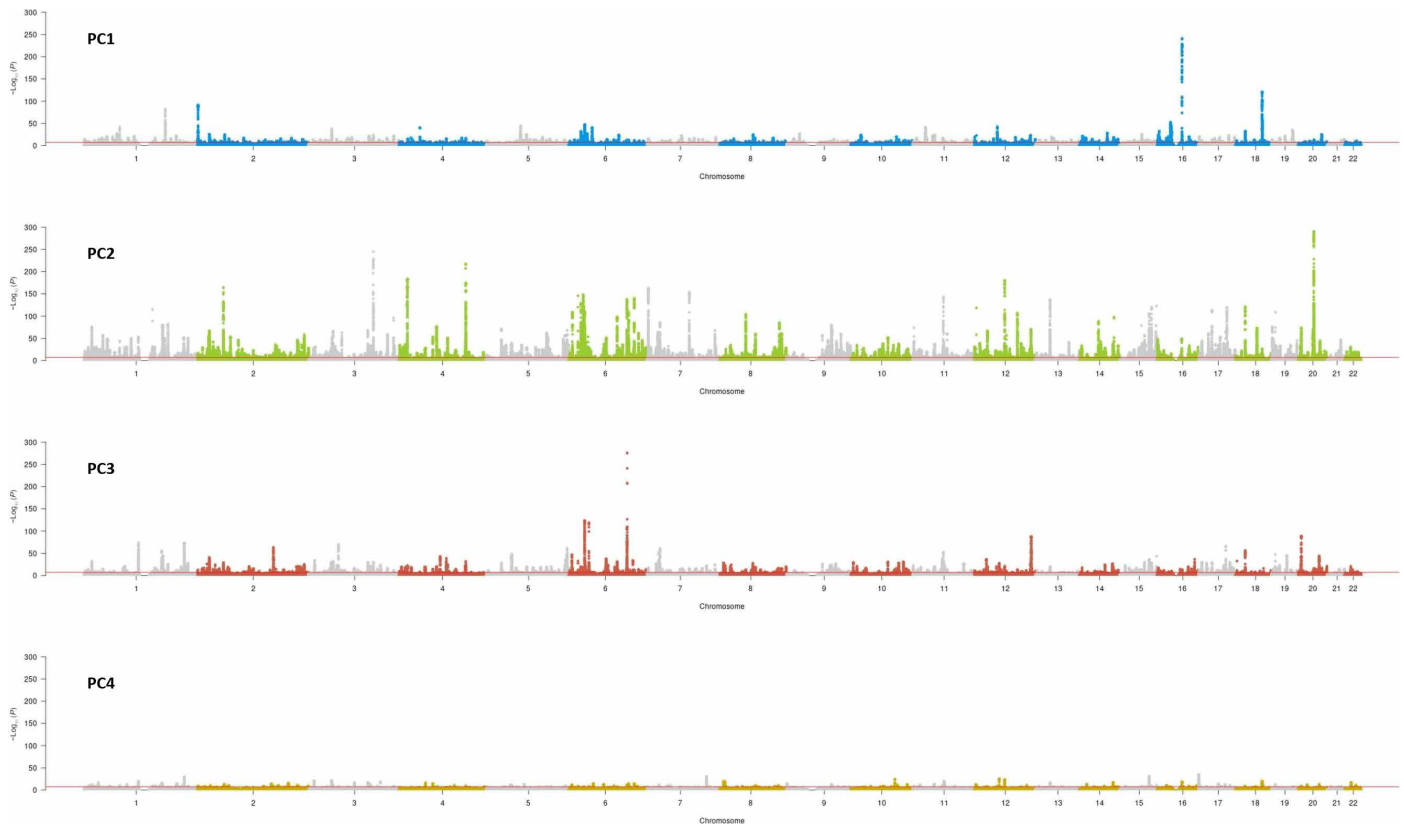


Fig. 3. Manhattan plots showing the genetic associations of the four body shapes.

brain and in the pituitary gland ($P < 7.67 \times 10^{-08}$); PC2 was enriched in the uterus, tibial nerve, ovary, cervix uteri, and blood vessels tissues ($P < 1.06 \times 10^{-3}$); PC3 showed enrichment in the adipose tissue, nerve, cervix uteri, uterus, blood vessel, breast, fallopian tube, and ovary ($P < 2.26 \times 10^{-4}$); and PC4 showed enrichment in uterus, blood vessel, adipose tissue, and muscle ($P < 2.13 \times 10^{-4}$). Overall, PC1 was enriched for gene expression corresponding to the neural broad transcription program (tissues with high proportion of neural cells), while PC2 to PC4 were enriched by gene expression corresponding to the mesenchymal broad transcription program (tissues with high proportion of mesenchymal cells) (21).

We further explored whether the genome-wide independent genetic variants were identified as expression quantitative trait loci (eQTL) or in LD ($r^2 > 0.8$) with an eQTL. For PC1, 204 and 110 variants of the 678 variants were identified as eQTLs in the brain tissues and the pituitary gland, respectively. Regarding the other PC GWAS results, several of the initially associated variants were also eQTLs of the tissues enriched for gene expression (PC2: 601 eQTLs in blood vessel, 592 in nerve, 144 in ovary, and 89 in uterus tissue; PC3: 278 eQTLs in adipose tissue, 259 in blood vessel, 154 in breast, 267 in nerve, 65 in ovary, and 40 in uterus tissue; and PC4: 83 eQTLs in adipose tissue, 79 in blood vessel, 68 in muscle, and 13 in uterus tissue). The eQTL with the lowest P value corresponded to rs72634819 for PC1 ($P = 1.7 \times 10^{-60}$), rs55938136 for PC2 ($P = 1.9 \times 10^{-184}$), rs34856835 for PC3 ($P = 2.2 \times 10^{-102}$), and rs10876864 for PC4 ($P = 1.1 \times 10^{-247}$), regulating the gene expression of *SLC35E2B*, *RPS26*, *XKR9*, and *LINC02210* genes, respectively.

MR of body shape phenotypes and CRC risk

Genetic variants related to body shape for the first three PCs identified by our GWAS formed strong instruments for the MR analysis (F statistic = 301.79 for PC1, 134.55 for PC2, and 52.22 for PC3) but not for PC4 (F statistic = 5.20 for PC4). Summary statistics for genetic associations with CRC risk for the 3414 genetic instruments of body shapes (672 for PC1, 1774 for PC2, 734 for PC3, and 193 for PC4) were obtained from a large CRC consortium [meta-analysis of Colorectal Transdisciplinary Study (CORECT), Colon Cancer Family Registry (CCFR), and Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)] (22).

PC1 was positively associated with overall CRC risk [inverse-variance weighted (IVW) odds ratio (OR) per 1 SD = 1.10; 95% CI: 1.07 to 1.13] and all subsites (colon, rectal, colon proximal, and colon distal) (Fig. 2). There was evidence of heterogeneity across genetic variants (Cochran's Q $P < 0.001$), with no evidence for directional pleiotropy as quantified using the MR Egger intercept test (table S5). Results in sensitivity analyses (MR Egger and weighted median approaches) supported the main IVW MR risk estimates (fig. S5).

PC3 was similarly positively associated with overall CRC risk (IVW OR = 1.12; 95% CI: 1.06 to 1.20), again consistent with the observational findings (Fig. 2) and sensitivity analyses (fig. S6). Heterogeneity was also detected ($P < 0.001$) (table S5). PC3 was positively associated with risk across all the CRC subsites and among women (IVW OR = 1.18; 95% CI: 1.10 to 1.28) (Fig. 2 and fig. S6).

MR estimates for PC2 showed a weak positive association with overall CRC but with a CI including the null (IVW OR = 1.03; 95%

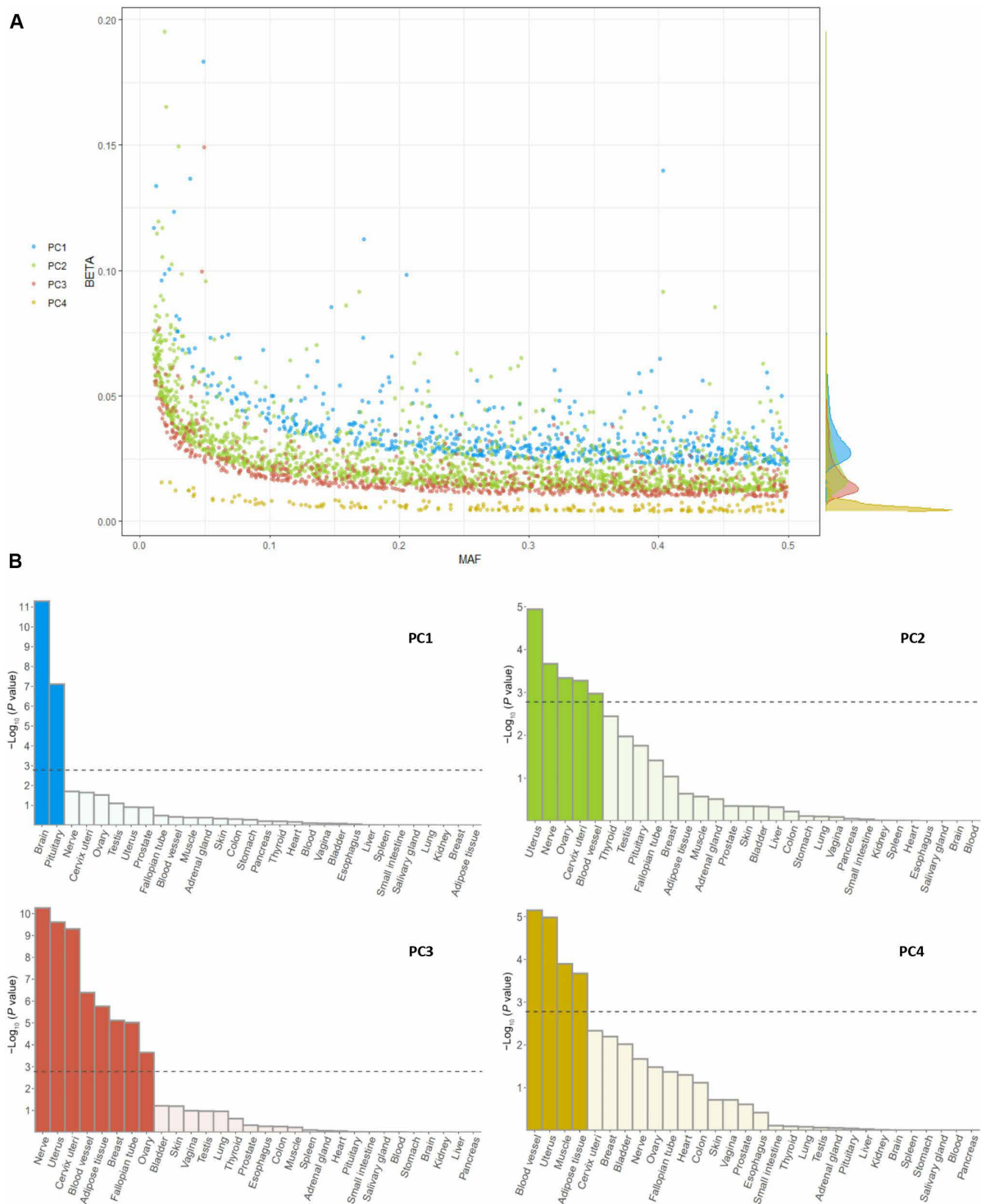


Fig. 4. Genetic architecture and tissue expression profile analyses of the four body shape phenotypes (PC1 to PC4). Genetic architecture (A) and tissue expression profile (B). Tissues reaching a P value of $<1.67 \times 10^{-3}$, after Bonferroni correction, are highlighted in darker colors. MAF, minor allele frequency.

CI: 1.00 to 1.06) (Fig. 2). Associations were similar among men and women, and CRC subsites (Fig. 2). However, neither the MR sensitivity analyses (fig. S7) nor the results obtained from the observational analysis supported these findings.

PC4 showed inconclusive associations with CRC risk (IVW OR = 1.05; 95% CI: 0.68 to 1.62) (Fig. 2 and fig. S8). The scatterplots depicting the genetic associations of the body shape-related instruments with the risk of overall CRC are shown in figs. S9 to S12.

MR of tissue-specific variants of body shapes and CRC

We next defined groups of tissue-specific genetic instruments for each of the four body shapes to assess whether these tissue-grouped associations with CRC risk show evidence of divergent mechanisms. Figure S13 shows the number of genetic instruments included in the MR analysis for each tissue and the number of genetic instruments shared between tissues.

The group of brain tissue-specific genetic instruments for PC1 (204 variants) was positively associated with overall CRC risk (IVW OR per 1 SD = 1.12; 95% CI: 1.05 to 1.19), while the PC1-pituitary gland expression instrument (110 variants) showed a more modest positive association (IVW OR = 1.07; 95% CI: 0.99 to 1.16) (Fig. 5). Sensitivity analyses supported these positive observations (fig. S14). The group of adipose tissue- and tibial nerve-genetic instruments for PC3 (278 and 267 variants, respectively) showed positive associations with overall CRC risk (IVW OR = 1.15; 95% CI: 1.05 to 1.27, IVW OR = 1.14; 95% CI: 1.03 to 1.27, respectively) (Fig. 5). Other tissue-gene expression instruments for PC3 showed less strong and/or robust associations with CRC risk (Fig. 5).

In an exploratory MR analysis of PC3 and CRC risk, we subdivided adipose tissue-specific genetic instruments and found no

substantial differences in risk estimates for CRC with ORs equal to 1.15 (95% CI: 1.00 to 1.34) and 1.13 (95% CI: 1.01 to 1.27) for visceral and subcutaneous adipose tissue, respectively. Further details on the heterogeneity and overall directional pleiotropy assessment can be found in table S6.

DISCUSSION

We conducted complementary observational and MR analyses to investigate the role of distinct body shapes in CRC development. We found that a generally obese body shape and a tall, centrally obese body shape were both robustly positively associated with overall CRC risk, its subsites (colon, distal colon, proximal colon, and rectal cancer), and in men and women. We also found that the generally obese body shape was associated with gene expression patterns in tissues with a high proportion of neural cells, whereas the tall and centrally obese body shape was associated with gene expression patterns in mesenchymal cells. This suggests that these two distinct body shapes may operate via separate downstream molecular pathways.

Traditionally, overall adiposity is quantified by BMI, while abdominal adiposity or body fat distribution is quantified by WC, HC, or WHR. However, BMI is correlated with height and body fat distribution measures (23), and the use of these indicators provides nonspecific risk estimates of the exposures under investigation (24, 25). To overcome this limitation, body shapes have been proposed to characterize different phenotypes of adiposity not fully captured by individual anthropometric traits. The body shapes in UK Biobank were extremely similar with those previously computed in the GIANT consortium (5), in the EPIC study (6), and between different ancestries available in UK Biobank (i.e., Caucasian, African,

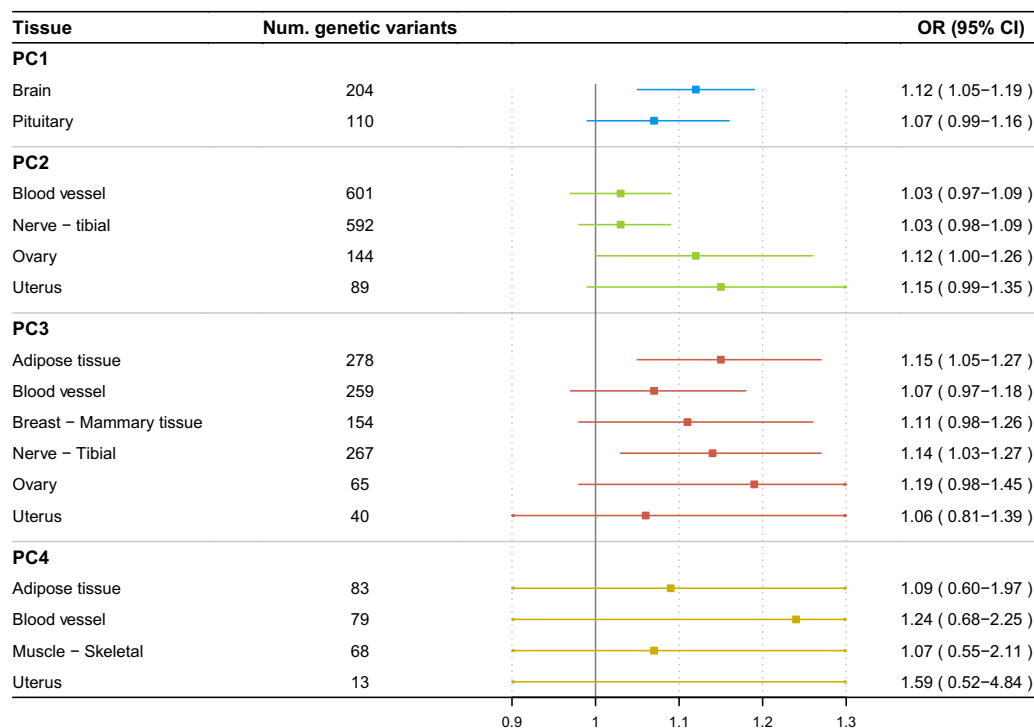


Fig. 5. Associations between body shape phenotype (PC1 to PC4) grouped gene sets and the risk of overall CRC. OR, odds ratio.

Asian, and Chinese). This emphasizes that body shape phenotypes can be built consistently across populations.

Although the body shapes may be more specific in assessing CRC risk in terms of adiposity and body size, the interpretability of the PCs is not straightforward. To facilitate the interpretation of the CRC risk estimates, we computed the means and the difference in SD units of each anthropometric trait for study participants in the top and bottom 5% of each body shape (table S7). For example, comparing top to bottom 5% of PC1 corresponded to almost uniform increments of 3.5 to 4.0 SDs for BMI, weight, WC, and HC, while increments in WHR and height were 1 and 0.4 SDs, respectively. In contrast, for PC3, we observed a 2-SD, 1-SD, and 0.3-SD increment in WHR, height, and WC, respectively, as well as decreases in HC (1.8 SD), BMI (1.3 SD), and weight (0.5 SD). Further detail and respective comparisons for PC2 and PC4 are shown in table S7.

To our knowledge, only one previous study, led by our group, investigated associations between these body shapes and the risk of cancer across 24 anatomical sites including colon and rectal cancers in an observational setting in EPIC (6). Our current findings are congruent and support potential causality of associations between body shapes and the risk of colon and rectal cancers observed in EPIC (6). Moreover, our study adds evidence for tissue-grouped causal associations of body shapes with CRC risk based on the tissue in which the body shape-mapped genes show markedly different expression levels.

Previous observational findings showed that excess body fat—measured by BMI—increased overall CRC risk by 5% per 1-SD increment (3), while an MR study estimated a 16% (per 1 SD) increased risk (26). Our findings for PC1 (generally obese body shape) showed more consistent estimates between observational and MR analyses, suggesting a risk increment per 1 SD of 13 and 10%, respectively. A meta-analysis of observational studies and an MR analysis, which used sex-specific genetic variants for BMI, showed stronger associations in men (8 and 23% increased risk) than in women (5 and 9% increased risk) (3, 26). Our results (both observational and MR) are congruent with a stronger association in men than women. We did not derive sex-specific genetic variants for our MR analysis because the loadings of the body shapes between women and men were very similar. This allowed for a larger sample size to identify the genetic instruments. The mean PC scores of the body shapes were also similar between men and women (table S8), except for slightly higher PC1 scores among men than women, suggesting that the prevalence of general obesity (PC1) was higher among men than women.

Our results for PC3 (tall, centrally obese) are probably best compared with previous observational and MR analyses that have linked abdominal adiposity (indicated by WHR) to CRC risk (3, 26). However, these previous studies differed in the estimates reporting 2% (per 1 SD) and 28% (per 1 SD) higher CRC risk for observational and MR analysis, respectively (3, 26). As in PC1, our estimates for both association studies were more consistent (9 and 12% increased risk per 1 SD). This could be explained by the specificity of body shape phenotype PC3 that also captures the contribution of other anthropometric traits such as height.

Also relevant is the finding that associations of PC2 (tall with a low WHR) with the risk of CRC were at best suggestive of a weak positive association based on the MR analysis, which, however, was not robust in MR sensitivity analyses and were conflicting with the observational analysis. This suggests that the association of height with CRC incidence might have been overestimated in previous

studies because of the difficulty to separate out the effect of height from WHR [WHR was not accounted for in previous studies (3, 27, 28)]. The differences with our results suggest that WHR may have an important role in the association between height and CRC risk, which is also supported by the results we obtained for PC3. Mechanistically, increased cancer risk due to height independently of (central) adiposity could be explained by an increased cell number in taller individuals (29). However, some biologic overlap with adiposity due to shared mechanistic pathways [e.g., elevated insulin-like growth factor 1 (IGF-1) levels] (30) cannot be ruled out and requires further investigation.

Viewed from a different mechanistic perspective with total bilirubin as a unifying indicator of metabolic health (31), PC2 was positively associated with total bilirubin in a UK Biobank study (31). This indicated a favorable metabolic health for individuals with higher PC2 scores [i.e., a lower trunk fat mass, lower circulating levels of triglycerides and of low-density lipoprotein (LDL)-cholesterol, and lower blood pressure]. In contrast, both PC1 and PC3, except for PC3 among women, were inversely associated with total bilirubin, indicating a less favorable metabolic health (31). Total bilirubin itself could potentially also play a causal role in CRC development, whereby higher circulating levels tended to be inversely associated with CRC risk in UK Biobank (32).

There is an intricate relationship between metabolic health, in particular dyslipidemia (i.e., imbalance of lipids such as cholesterol and triglycerides), and inflammation (33), which could present a mechanistic link to CRC. However, it is thought that inflammation leads to changes in lipid metabolism rather than the other way round; for example, tumor necrosis factor- α (TNF- α) stimulates triglyceride synthesis in hepatic cell lines (33). It is also likely that statins (which is a drug prescribed to reduce LDL-cholesterol levels) could confound the dyslipidemia-CRC associations. Findings from prospective and MR studies of either dyslipidemia or statins with CRC risk are inconsistent in this respect (34–37). Collectively, it remains to be clarified whether dyslipidemia plays a causal role in CRC development or is only a bystander.

Our GWAS identified 3414 independent genetic variants that were robustly associated with at least one of the four body shapes and expands prior knowledge. Ried *et al.* (5) had identified 207 independent genetic variants, of which 92.75% were replicated in our GWAS, which clearly shows the importance of our study with an augmented sample size [$N = 460,198$ current study versus $N = 173,278$ in Ried *et al.* (5)]. A total of 59 genetic variants were common with at least two of the body shapes. Overall, our results suggest different genetic architectures among body shapes.

In our tissue expression enrichment analyses, we found that the genetic variants related to PC1 showed gene expression enrichment in tissues with high proportion of neural cells (brain and pituitary gland tissues), while genetic variants related to PC3 showed gene expression enrichment in tissues with high proportion of mesenchymal cells (adipose, breast, nerve, blood vessel tissues, and the reproductive tissues of uterus, cervix, ovary, and fallopian tube) (21). Overall, the observed difference in genetic and tissue expressions suggests that PC1 and PC3 may capture different molecular origins and metabolic consequences of these body shapes and may reflect divergent mechanisms by which body shapes influence the risk of CRC. One mechanism by which PC1-influencing genetic variants expressed in brain tissue likely exert their effects is via energy homeostasis, which is strongly controlled by the central nervous system (38). Conforming

with this notion, genetic variants strongly associated with PC1 were found in the *AKT3* and *FTO* genes, which are known to be involved in insulin signaling and energy homeostasis, respectively (13, 14). Perturbed energy homeostasis can lead to neurogenic adiposity causing chronic, systemic inflammation mainly through the overexpression of proinflammatory adipokines (39). In contrast, PC3-influencing genetic variants expressed in adipose tissue might exert their effects through an unfavorable fat distribution characterized by less subcutaneous fat storage and/or less fat accumulation at legs and hips (40). This could lead to metabolically harmful ectopic fat storage (e.g., in liver and muscle) and may result in increased insulin resistance (40). In addition to nerves and adipose tissue, PC3 also showed enrichment in female-specific reproductive tissues such as the uterus. Whether this could explain the heterogeneity between men and women in the MR estimates for PC3 and CRC risk needs further study.

The main strength of this study is the combination of prospective observational and instrumental analyses using two large independent datasets: UK Biobank for the observational and GWAS analyses, and the CRC genetic consortia dataset for the MR approach. Second, we identified strong genetic instruments in the GWAS analysis (except for PC4), which allowed us to reliably estimate CRC risk in our MR analysis. Third, we performed multiple sensitivity analyses to evaluate the influence of residual confounding by smoking and reverse causation bias in the observational setting and to assess the potential violation of the MR assumptions.

Our findings should be interpreted considering the following limitations. Despite adjusting for the main confounders, residual confounding can be present in the risk estimates of prospective observational analysis. Second, a one-point-in-time assessment of body shapes may underestimate cumulative exposure over the life course or miss changes in body shapes. Other limitations include that, in general, the MR studies can be slightly biased because of uncontrolled confounding from family effects such as assortative mating, dynastic effects, and population structure (41). We minimized this bias from population stratification performing a two-sample MR analysis in samples composed of European ancestry individuals. Last, although our tissue-specific MR results suggest divergent mechanisms across tissues, we acknowledge that the observed differences in risk across tissues could be influenced by varying instrument strength/power. Future studies could use genetic colocalization to verify the functional relationship between the body shape PCs and gene expression.

In these complementary observational and MR analyses, we found that a generally obese and a distinctly tall, centrally obese body shape were both robustly associated with a higher CRC risk. While the former was enriched for gene expression in the brain and pituitary gland, the latter was enriched in adipose, nerve, blood vessel, breast, and reproductive tissues. Together with the results of the tissue-specific MR analyses, our work suggests that these two distinct body shapes may operate via separate molecular pathways through which they influence the risk of CRC. Future studies that examined these two pathways in mechanistic/experimental studies would be welcomed.

MATERIALS AND METHODS

Ethics statement

All participants from the UK Biobank have provided written informed consent to participate. UK Biobank has ethical approval from the Northwest Multi-Centre Research Ethics Committee. All

participants from the GECCO, CORECT, and CCFR consortium provided written informed consent, and each study was approved by the relevant research ethics committee or institutional research board. GECCO is approved under Fred Hutch Cancer Center IRB file no. 3995.

UK Biobank

UK Biobank is a prospective cohort study with over 500,000 individuals with an age range of 38 to 73 years at recruitment between 2006 and 2010 from 22 study centers in England, Scotland, and Wales. Data on anthropometry, lifestyle, health-related indicators, and biological samples were collected (42, 43). Up to 488,377 individuals were genotyped using either a custom Affymetrix UK Biobank Axiom Array chip or a custom Affymetrix UK BiLEVE Axiom Array chip (43).

Body shape phenotypes

Anthropometric measures were recorded by trained personnel. Height and weight were measured using a Seca 202 stadiometer and a Tanita BC-418 body composition analyzer, respectively. BMI was calculated as body weight (in kilograms) divided by height (in meters squared). WC was measured at the natural indent or the umbilicus and HC at the widest point of the participants. WHR was calculated as WC (in centimeters) divided by HC (in centimeters) (42). Weight and WC were not recorded for women, who were pregnant at baseline, and were thus excluded in our analysis because of missing exposure information.

Body shapes were derived from a PCA on the residuals of the six anthropometric traits (BMI, height, weight, WHR, WC, and HC) computed by regressing each trait on sex, age, and recruitment center. The distributions of PC data were winsorized at 1 and 99% to reduce the influence of point anomalies (44). Details regarding PC loadings and its proportion of variance explained can be found in table S1. We computed the arithmetic mean of each anthropometric trait of participants in the top and bottom 5% across all four body shape phenotypes to facilitate its interpretation (table S7). We also derived the body shape phenotypes in other ancestries available in UK Biobank (i.e., African, Asian, and Chinese).

Genome-wide association study for body shape phenotypes

We excluded participants for excess of heterozygosity, extensive missingness (rate > 5%), sex discrepancy, aneuploidy, and discordance across control replicates ($N = 1865$). We restricted our analysis to European ancestry individuals based on self-reported ancestry and categories derived using *k*-means algorithm to UK Biobank genetic PCs (fig. S15). Last, we also excluded individuals showing nonadherence to follow-up appointments, obtaining a final sample size of 460,198 European individuals. A total of 8,889,767 genetic variants remained for the GWAS after applying filters in missingness < 0.05, minor allele frequency (MAF) > 0.01, imputation info-score > 0.1, and Hardy-Weinberg equilibrium (HWE) test $P > 1 \times 10^{-8}$. These quality control steps were performed using PLINK2 except for the *k*-means algorithm, which was implemented using the “stats” package in the statistical software R 4.2.1 and RStudio 2022.7.0.548 (45–48).

Genome-wide association testing was implemented using Bayesian linear mixed-model analysis BOLT-LMM v2.3.5 (49). BOLT-LMM requires three primary components to be performed: the imputed genetic data for the association testing, a reference panel of LD scores per genetic variant, and genotype data used to approximate

a genetic relationship matrix (GRM). As our sample was restricted to European ancestry individuals, we were able to use the LD score reference panel provided by BOLT-LMM, which is based on 1000 Genomes samples using the LD Score Regression (LDSC) software (50). The genetic variant selection for the GRM was as follows: missingness < 0.01 , HWE $< 1 \times 10^{-08}$, MAF > 0.01 , and imputation info-score > 0.8 . The remaining genetic variants were hard called using a threshold of 0.25 and pruned at $r^2 < 0.2$. Some conflictive genomic regions such as the lactase locus on chromosome 2, the major histocompatibility complex (MHC) on chromosome 6, and inversions on chromosomes 8 and 17 were removed. Last, the GRM included 836,810 genetic variants.

BOLT-LMM was applied using the standard infinitesimal model adjusting for age, age², sex, genotyping batch, and 10 genetic PCs. To identify independent genetic variants associated with each body shape, we performed an LD clumping on genetic variants with a P value of $< 5 \times 10^{-08}$ in windows of 10 Mb, with a clumping threshold at $r^2 = 0.01$, using the “ieugwasr” package in R 4.2.1 and RStudio 2022.7.0.548 (47, 48, 51, 52).

Genetic variants—not previously linked to BMI, weight, height, WHR, WC, and HC—were defined on the basis of databases from the NHGRI-EBI GWAS Catalog, the MRC IEU OpenGWAS project, the Open Targets Genetics, and PhenoScanner (accessed in February 2023) (15–19, 52, 53). In addition, we also compared our results with the latest GWAS on height (sample size of 5.4 million individuals) (54).

Tissue expression enrichment analysis

A tissue expression enrichment analysis on the GWAS summary statistics of each of the four body shapes was carried out using the MAGMA gene set analysis implemented in the Functional Mapping and Annotation (FUMA) of GWAS software (55, 56). This analysis was performed for 30 general tissue types (e.g., brain tissue) and separately for 54 tissue types that included specific subtypes (e.g., 13 brain subregions). Some tissues had no subtypes (e.g., lung) and were included in both analyses. In the first step, among all genetic variants of the GWAS on body shapes, only the genetic variants mapped to a protein-coding gene were selected. Second, a regression model was fitted using the normalized GWAS-derived P values of the genetic variants identified in the first step as the outcome and the average gene expression per tissue type and several technical confounders as covariables. The tissues were derived from a total of 838 postmortem donors of the GTEx V8 dataset (57). The sample numbers from genotyped donors across the 54 tissue types are given in table S9.

Genetic variants related to body shape phenotypes

From the identified independent genetic variants across the four body shapes, we estimated the variance explained by the genetic variants related to each body shape, as a function of the effect size (β) for the phenotype in SD units and the MAF of the variants included [cumulative of $2 \times \beta^2 \times \text{MAF} \times (1 - \text{MAF})$] (58). We also calculated the F statistic to quantify the strength of the relationship between each genetic instrument and its body shape. The F statistic is an estimation of the magnitude of the instrument bias (with an F statistic of < 10 indicating weak instruments) (59).

For the tissue-specific MR analysis, we selected genetic instruments that were found in the GWAS and identified as an eQTL or in LD ($r^2 > 0.8$) with an eQTL through MAGMA gene set analysis. We

restricted our tissue-specific MR analyses to the 30 general tissues using as genetic instruments all genetic variants identified for each tissue subtype (e.g., genetic instruments for adipose tissue included genetic instruments mapped to subcutaneous and visceral adipose tissues). Further details of the number of genetic variants mapped in each specific tissue and the different specific tissues composing the general ones can be found in table S10.

Genetic variants related to CRC risk

Summary statistics for CRC risk of the genetic variants related to body shapes were obtained from three large CRC genetic consortia: GECCO, CORECT, and CCFR. These datasets did not include UK Biobank participants to avoid sample overlap between exposures and outcomes. The dataset consisted of 98,715 participants of European ancestry (52,775 cases and 45,940 controls). Further analyses by subsite [colon (27,817 cases), rectal (13,713 cases), distal colon (14,016 cases), and proximal colon (12,360 cases) cancer] and by sex (24,568 women and 28,207 men cases) were performed (22).

MR analysis

MR analysis is an approach that uses genetic variants as instruments to seek potential causal associations between an exposure and an outcome. MR is based on three main assumptions: (i) The genetic variants are strongly associated with the exposure, (ii) the genetic variants are not associated with any potential confounder, and (iii) the genetic variants are associated with the outcome only through the exposure (60–62).

Two-sample summary-based MR approach consists of the application of MR methods to summary-level data of a GWAS on the exposures (the four body shapes) and a GWAS on the outcomes (overall CRC, by subsites, and by sex) (63). Random-effects IVW MR method was used as the primary analysis (64).

We used scatterplots to present the genetic associations between body shapes and CRC risk to visually examine the consistency of MR estimates and the potential associated bias.

Last, we also performed the same MR analysis design for the tissue-specific MR approach.

Sensitivity analyses

The first MR assumption (i.e., the genetic variants are associated with the exposure) is likely to be met as we selected genetic variants with a rigorous P value of $< 5 \times 10^{-8}$. The second (i.e., the genetic variants are not associated with any potential confounder) and third (i.e., the genetic variants are associated with the outcome only through the exposure) MR assumptions are more difficult to test (65). Thus, several sensitivity analyses were implemented to assess the potential violation of these assumptions.

We used the Cochran's Q test to evaluate heterogeneity among causal effects derived from each of the selected variants, considering the existence of heterogeneity with a P value of < 0.05 (66). MR-Egger regression was also implemented providing a statistical test for overall directional pleiotropy (intercept term different from zero) (67). We performed the weighted median approach as it allows for the violations of the second and the third assumptions when up to 50% of the genetic variants are invalid (i.e., violation of one or more MR assumptions) (68, 69). Both tests provide valid MR estimates in the presence of overall directional pleiotropy but suffer from reduced power (64, 67). All analyses were conducted using the “MendelianRandomization” in R 4.2.1 and RStudio 2022.7.0.548 (47, 48, 70).

Observational analysis**CRC incident cases**

Data on cancer diagnoses were provided by National Health Service (NHS) Digital and Public Health England for participants from England and Wales and by NHS Central Register (NHSCR) for participants residing in Scotland. Detailed information about the record linkage and cancer registry information can be found at <https://biobank.ndph.ox.ac.uk/~bbdata/CancerNumbersReport.html>. Type of cancer cases was coded using the International Classification of Diseases (ICD), and information on tumor morphology and histology was derived using the third revision of the International Classification of Diseases for Oncology (71). Date of complete follow-up was 29 February 2020 for England and Wales and 31 January 2021 for Scotland. CRC was defined as ICD C18-C20 (3728 cases), colon cancer as C18 (2443 cases), and rectal cancer as C19-C20 (1285 cases) using the 10th Edition of the International Classification of Diseases (ICD-10). Further subdivided, proximal colon cancer was defined as C18.0-C18.5 (1370 cases) and distal colon cancer as C18.6-C18.7 (966 cases).

Covariables

We identified potential confounding variables using directed acyclic graphs (Fig. 6 and fig. S16) (72, 73). These were socioeconomic factors: age, sex, recruitment center, Townsend deprivation index, and education (university/college degree, A levels/AS levels/National Vocational Qualification/Higher National Diploma/Higher National Certificate or equivalent/other professional qualification, O levels/Certificate of Secondary Education or equivalent, none of the above); ethnicity (white, mixed, Asian/British Asian, Black/Black British, Chinese, other); lifestyle factors: tobacco smoking (never, previous, current), physical activity (metabolic equivalent-minutes per week), sedentary behavior (hours per day spent watching television/using computer/driving), adherence to a healthy diet score based on several food items (processed meat, red meat, oily fish, milk type, spread type, cereal intake, and fruit and vegetables) (74), milk intake (<150, 150 to 299, ≥300 ml/day), and alcohol intake frequency (daily or almost daily, three or four times a week, once or twice a week, one to three times a month, special occasions only, never); medication

use: NSAIDs and hormone therapy in postmenopausal women; bowel (CRC) cancer screening and family history of CRC (father and/or mother). Further details on these variables are provided in the UK Biobank protocol (<http://ukbiobank.ac.uk>).

We did not adjust for comorbidities, such as type 2 diabetes, because we assumed that this comorbidity would rather be a mediator (i.e., in the pathway from body shapes to CRC) than a confounder. Unmeasured (known) confounders are inflammatory bowel syndrome and Lynch syndrome; both phenotypes are difficult to diagnose clinically, and data availability in UK Biobank was therefore limited. However, as indicated in the DAG, we assumed that the confounding paths for both phenotypes are at least partly blocked by accounting for bowel (CRC) cancer screening and family history of CRC. After the exclusion of participants with prevalent cancer at baseline, and missing, incomplete, or implausible information on exposures or in at least one covariable, the sample for observational analyses comprised 329,828 participants (51.1% women; fig. S17).

Statistical analysis

Cox proportional hazards models with age as the underlying timescale were used. We estimated HRs and 95% CIs per 1-SD increment of each PC. Models were stratified by sex, age (in 5-year categories), and recruitment center and adjusted for covariables (see above) in the main models. Proportional hazards assumptions were tested with scaled Schoenfeld residuals (75), and log-likelihood ratio tests were used to assess potential departure from linearity for all exposure variables.

Sensitivity analyses

As sensitivity analyses, we excluded the current and former smokers to address potential residual confounding by smoking, excluded the two initial years of follow-up to control for reverse causation, and additionally fitted a crude model that included all four PCs.

All statistical tests were two-sided and P values < 0.05 were considered statistically significant. All analyses were performed with R version 4.2.3 using the packages “FactoMineR” and “rms” (47, 76, 77).

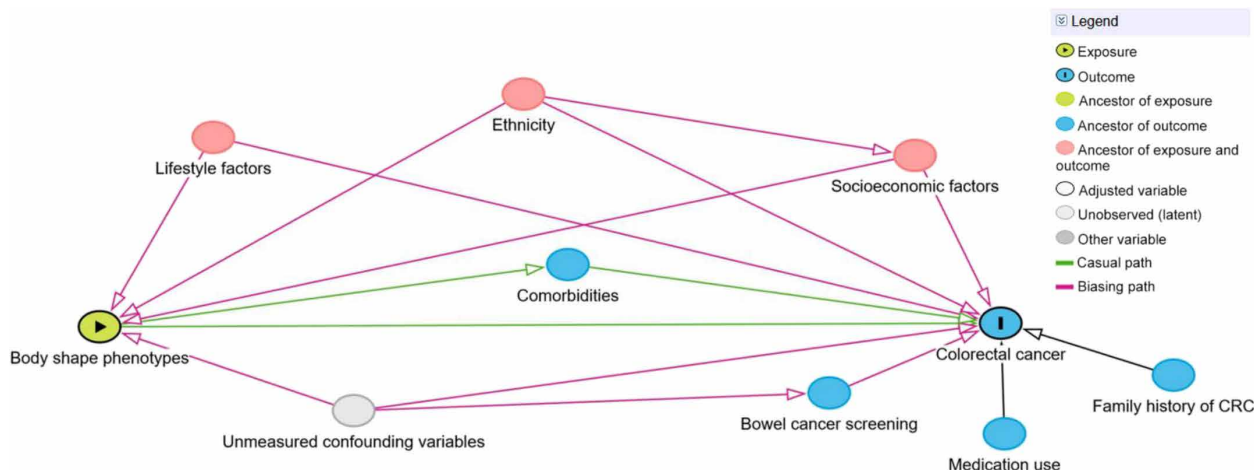


Fig. 6. Directed acyclic graph depicting the assumed causal relationship between body shape phenotypes and CRC risk with its confounding and mediating paths.

Supplementary Materials

This PDF file includes:

Figs. S1 to S17

Tables S1 to S10

Supplementary Text

Other Supplementary Material for this manuscript includes the following:

Data S1

REFERENCES AND NOTES

- H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, F. Bray, Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **149**, 778–789 (2021).
- World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report, Diet, nutrition, physical activity and colorectal cancer (2018).
- H. Freisling, M. Arnold, I. Soerjomataram, M. G. O'Doherty, J. M. Ordóñez-Mena, C. Bamia, E. Kampman, M. Leitzmann, I. Romieu, F. Kee, K. Tsilidis, A. Tjønneland, A. Trichopoulos, P. Boffetta, V. Benetou, H. B. Bueno-De-Mesquita, J. M. Huerta, H. Brenner, T. Wilsgaard, M. Jenab, Comparison of general obesity and measures of body fat distribution in older adults in relation to cancer risk: Meta-analysis of individual participant data of seven prospective cohorts in Europe. *Br. J. Cancer* **116**, 1486–1497 (2017).
- J. S. Ried, J. M. Jeff, M. A. Y. Chu, J. L. Bragg-Gresham, J. van Dongen, J. E. Huffman, T. S. Ahluwalia, G. Cadby, N. Eklund, J. Eriksson, T. Esko, M. F. Feitosa, A. Goel, M. Gorski, C. Hayward, N. L. Heard-Costa, A. U. Jackson, E. J. J. Kinnunen, S. Kanoni, K. Kristiansson, Z. Kutalik, J. Lahti, J. Luan, R. Mägi, A. Mahajan, M. Mangino, C. Medina-Gomez, K. L. Monda, I. M. Nolte, L. Pérusse, I. Prokopenko, L. Qi, L. M. Rose, E. Salvi, M. T. Smith, H. Snieder, A. Stancáková, Y. J. Sung, I. Tachmazidou, A. Teumer, G. Thorleifsson, P. van der Harst, R. W. Walker, S. R. Wang, S. H. Wild, S. M. Willems, A. Wong, W. Zhang, E. Albrecht, A. C. Alves, S. J. L. Bakker, C. Barlassina, T. M. Bartz, J. Beilby, C. Bellis, R. N. Bergman, S. Bergmann, J. Blangero, M. Blüher, E. Boerwinkle, L. L. Bonnycastle, S. R. Bornstein, M. Bruinenberg, H. Campbell, Y.-D. I. Chen, C. W. K. Chiang, P. S. Chines, F. S. Collins, F. Cucca, L. A. Cupples, F. D'Ávila, E. J. C. de Geus, G. Dedoussis, M. Dimitriou, A. Döring, J. G. Eriksson, A.-E. Farmaki, M. Farrall, T. Ferreira, K. Fischer, N. G. Forouhi, N. Friedrich, A. P. Gjesing, N. Glorio, M. Graff, H. Grallert, N. Grarup, J. Gräßler, J. Grewal, A. Hamsten, M. N. Harder, C. A. Hartman, M. Hassinen, N. Hastie, A. T. Hattersley, A. S. Havulinna, M. Heliövaara, H. Hillege, A. Hofman, O. Holmen, G. Homuth, J.-J. Hottenga, J. Hui, L. L. Husemoen, P. G. Hysi, A. Isaacs, T. Ittermann, S. Jalilzadeh, A. L. James, T. Jørgensen, P. Jousilahti, A. Jula, J. M. Justesen, A. E. Justice, M. Kähönen, M. Karaleftheri, K. T. Khaw, S. M. Keinanen-Kiukkaanniemi, L. Kinnunen, P. B. Knekt, H. A. Koistinen, I. Kolcic, I. K. Kooper, S. Koskinen, P. Kovacs, T. Kyriakou, T. Laitinen, C. Langenberg, A. M. Lewin, P. Lichtner, C. M. Lindgren, J. Lindström, A. Linneberg, R. Lorbeer, M. Lorentzon, R. Luben, V. Lyssenko, S. Männistö, P. Manunta, I. M. Leach, W. L. M. Ardle, B. McKnight, K. L. Mohlke, E. Mihailov, L. Milani, R. Mills, M. E. Montasser, A. P. Morris, G. Müller, A. W. Musk, N. Narisu, K. K. Ong, B. A. Oostra, C. Osmond, A. Palotie, J. S. Pankow, L. Paternoster, B. W. Penninx, I. Pichler, M. G. Pilia, O. Polasek, P. P. Pramstaller, O. T. Raitakari, T. Rankinen, D. C. Rao, N. W. Rayner, R. Ribel-Madsen, T. K. Rice, M. Richards, P. M. Ridker, F. Rivadeneira, K. A. Ryan, S. Sanna, M. A. Sarzynski, S. Scholtens, R. A. Scott, S. Sebert, L. Southam, T. H. Sparsø, V. Steinthorsdottir, K. Stirrups, R. P. Stolk, K. Strauch, H. M. Stringham, M. A. Swertz, A. J. Swift, A. Tönjes, E. Tsafantakis, P. J. van der Most, J. V. Van Vliet-Ostaptchouk, L. Vandenput, E. Vartiainen, C. Venturini, N. Verweij, J. S. Viikari, V. Vitart, M.-C. Vohl, J. M. Vonk, G. Waeber, E. Widén, G. Willemsen, T. Wilsgaard, T. W. Winkler, A. F. Wright, L. M. Yerges-Armstrong, J. H. Zhao, M. C. Zillikens, D. I. Boomsma, C. Bouchard, J. C. Chambers, D. I. Chasman, D. Cusi, R. T. Gansevoort, C. Gieger, T. Hansen, A. A. Hicks, F. Hu, K. Hveim, M.-R. Jarvelin, E. Kajantie, J. S. Kooner, D. Kuh, J. Kuusisto, M. Laakso, T. A. Lakka, T. Lehtimäki, A. Metspalu, I. Njølstad, C. Ohlsson, A. J. Oldehinkel, L. J. Palmer, O. Pedersen, M. Perola, A. Peters, B. M. Psaty, H. Puolijoki, R. Rauramaa, I. Rudan, V. Salomaa, P. E. H. Schwarz, A. R. Shudiner, J. H. Smit, T. I. A. Sørensen, T. D. Spector, K. Stefansson, M. Stumvoll, A. Tremblay, J. Tuomilehto, A. G. Uitterlinden, M. Uusitupa, U. Völker, P. Vollenweider, N. J. Wareham, H. Watkins, J. F. Wilson, E. Zeggini, G. R. Abecasis, M. Boehnke, I. B. Borecki, P. Deloukas, C. M. van Duijn, C. Fox, L. C. Groop, I. M. Heid, D. J. Hunter, R. C. Kaplan, M. I. M. Carthy, K. E. North, J. R. O'Connell, D. Schlessinger, U. Thorsteinsdottir, D. P. Strachan, T. Frayling, J. N. Hirschhorn, M. Müller-Nurasyid, R. J. F. Loos, A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. *Nat. Commun.* **7**, 13357 (2016).
- A. M. Sedlmeier, V. Viallon, P. Ferrari, L. Peruchet-Noray, E. Fontvieille, A. Amadou, N. Seyed Khoei, A. Weber, H. Baurecht, A. K. Heath, K. Tsilidis, R. Kaaks, V. Katzke, E. Inan-Eroglu, M. B. Schulze, K. Overvad, C. Bonet, E. Ubago-Guisado, M.-D. Chirlaque, E. Ardanaz, A. Perez-Cornago, V. Pala, R. Tumino, C. Sacerdote, F. Pisanis, K. B. Borch, C. Rylander, E. Weiderpass, M. J. Gunter, B. Fervers, M. F. Leitzmann, H. Freisling, Body shape phenotypes of multiple anthropometric traits and cancer risk: A multi-national cohort study. *Br. J. Cancer* **128**, 594–605 (2022).
- L. H. Nguyen, A. Goel, D. C. Chung, Pathways of colorectal carcinogenesis. *Gastroenterology* **158**, 291–302 (2020).
- G. D. Smith, G. Hemani, Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–R98 (2014).
- D. A. Lawlor, K. Tilling, G. D. Smith, Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).
- A. E. Locke, B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers, F. R. Day, C. Powell, S. Vedantam, M. L. Buchkovich, J. Yang, D. C. Croteau-Chonka, T. Esko, T. Fall, T. Ferreira, S. Gustafsson, Z. Kutalik, J. Luan, R. Mägi, J. C. Randall, T. W. Winkler, A. R. Wood, T. Workalemahu, J. D. Faul, J. A. Smith, J. H. Zhao, W. Zhao, J. Chen, R. Fehrmann, Å. K. Hedman, J. Karjalainen, E. M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J. L. Bolton, J. L. Bragg-Gresham, S. Buyske, A. Demirkan, G. Deng, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, A. Goel, J. Gong, A. U. Jackson, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, M. Mangino, I. M. Leach, C. Medina-Gomez, S. E. Medland, M. A. Nalls, C. D. Palmer, D. Pasko, S. Pechlivanis, M. J. Peters, I. Prokopenko, D. Shungin, L. L. Stančáková, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, A. Isaacs, E. Albrecht, J. Ärnlöv, G. M. Arscott, A. P. Attwood, S. Bandinelli, A. Barrett, I. N. Bas, C. Bellis, A. J. Bennett, C. Berne, R. Blagieva, M. Blüher, S. Böhringer, L. L. Bonnycastle, Y. Böttcher, H. A. Boyd, M. Bruinenberg, I. H. Caspersen, Y.-D. I. Chen, R. Clarke, E. W. Daw, A. J. M. de Craen, G. Delgado, M. Dimitriou, A. S. F. Doney, N. Eklund, K. Estrada, E. Eury, L. Folkersen, R. M. Fraser, M. E. Garcia, F. Geller, V. Giedraitis, B. Gigante, A. S. Go, A. Golay, A. H. Goodall, S. D. Gordon, M. Gorski, H.-J. Grabe, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, C. J. Groves, G. Gusto, J. Haessler, P. Hall, T. Haller, G. Hallmans, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, C. Hengstenberg, O. Holmen, J.-J. Hottenga, A. L. James, J. M. Jeff, Å. Johansson, J. Jolley, T. Juliusdottir, L. Kinnunen, W. Koenig, M. Koskenvuo, W. Kratzer, J. Laitinen, K. Lamina, K. Leander, N. R. Lee, P. Lichtner, L. Lind, J. Lindström, K. S. Lo, S. Lobbens, R. Lorbeer, Y. Lu, F. Mach, P. K. E. Magnusson, A. Mahajan, W. L. McArdle, S. M. Lachlan, C. Menni, S. Merger, E. Mihailov, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, A. Mulas, G. Müller, M. Müller-Nurasyid, A. W. Musk, R. Nagaraja, M. M. Nöthen, I. M. Nolte, S. Pilz, N. W. Rayner, F. Renstrom, R. Rettig, J. S. Ried, S. Ripke, N. R. Robertson, L. M. Rose, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, W. R. Scott, T. Seufferlein, J. Shi, A. V. Smith, J. Smolonska, A. V. Stanton, V. Steinthorsdottir, K. Stirrups, H. M. Stringham, J. Sundström, M. A. Swertz, A. J. Swift, A.-C. Syvänen, S.-T. Tan, B. O. Tayo, B. Thorand, G. Thorleifsson, J. P. Tyrer, H.-W. Uh, L. Vandenput, F. C. Verhulst, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, H. R. Warren, D. Waterworth, M. N. Weedon, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, The LifeLines Cohort Study, E. P. Brennan, M. Choi, Z. Dastani, A. W. Drong, P. Eriksson, A. Franco-Cereceda, J. R. Gadin, A. G. Gharavi, M. E. Goddard, R. E. Handsaker, J. Huang, F. Karpe, S. Kathiresan, S. Keildson, K. Kiryluk, M. Kubo, J.-Y. Lee, L. Liang, R. P. Lifton, B. Ma, S. A. M. Carroll, A. J. M. Knight, J. L. Min, M. F. Moffatt, G. W. Montgomery, J. M. Murabito, G. Nicholson, D. R. Nyholt, Y. Okada, J. R. B. Perry, R. Dorajoo, E. Reinmaa, R. M. Salem, N. Sandholm, R. A. Scott, L. Stolk, A. Takahashi, T. Tanaka, F. M. van 't Hooft, A. A. E. Vinkhuyzen, H.-J. Westra, W. Zheng, K. T. Zondervan, The ADIPOGen Consortium; The AGEN-BMI Working Group; The CARDIOGRAMplusC4D Consortium; The CKDGen Consortium; The GLGC; The ICBP; The MAGIC Investigators; The MuTHER Consortium; The MiGen Consortium; The PAGE Consortium; The ReproGen Consortium; The GENIE Consortium; The International Endogene Consortium; A. C. Heath, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, J. Blangero, P. Bovet, H. Campbell, M. J. Caulfield, G. Cesana, A. Chakravarti, D. I. Chasman, P. S. Chines, F. S. Collins, D. C. Crawford, L. A. Cupples, D. Cusi, J. Danesh, U. de Faire, H. M. den Ruijter, A. F. Dominiczak, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, S. B. Felix, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, O. H. Franco, R. T. Gansevoort, P. V. Gejman, C. Gieger, O. Gottesman, V. Gudnason, U. Gyllenstein, A. S. Hall, T. B. Harris, A. A. Hattersley, A. A. Hicks, L. A. Hindorf, A. D. Hingorani, A. Hofman, G. Homuth, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hyppönen, T. Illig, K. B. Jacobs, M.-R. Jarvelin, K.-H. Jöckel, B. Johansen, P. Jousilahti, J. W. Jukema, A. M. Jula, J. Kaprio, J. J. P. Kastelein, S. M. Keinanen-Kiukkaanniemi, L. A. Kiemeny, P. Knekt, J. S. Kooner, C. Kooperberg, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. L. Marchand, T. Lehtimäki, V. Lyssenko, S. Männistö, A. Marette, T. C. Matise, C. A. McKenzie, B. M. Knight, F. L. Moll, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, P. A. F. Madden, G. Pasterkamp, J. F. Peden, A. Peters, D. S. Postma, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, P. M. Ridker, J. D. Ríoum, M. D. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, H. Schunkert, P. E. H. Schwarz, P. Sever, A. R. Shuldiner, J. Sinisalo, R. P. Stolk, K. Strauch, A. Tönjes, D.-A. Trégouët, A. Tremblay, E. Tremoli, J. Virtamo, M.-C. Vohl, U. Völker, G. Waeber, G. Willemsen, J. C. Witterman, M. C. Zillikens, L. S. Adair, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, S. R. Bornstein, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper,

- P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, J. Hui, D. J. Hunter, K. Hveem, R. C. Kaplan, M. Kivimäki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, A. Metspalu, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T. E. Saaristo, D. Saleheen, N. Sattar, E. E. Schadt, D. Schlesinger, P. E. Slagboom, H. Snieder, T. D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A. G. Uitterlinden, M. Usutupa, P. van der Harst, M. Walker, H. Wallaschofski, N. J. Wareham, H. Watkins, D. R. Weir, H.-E. Wichmann, J. F. Wilson, P. Zanen, I. B. Borecki, P. Deloukas, C. S. Fox, I. M. Heid, J. R. O'Connell, D. P. Strachan, K. Stefansson, C. M. van Duijn, G. R. Abecasis, L. Franke, T. M. Frayling, M. I. McCarthy, P. M. Visscher, A. Scherag, C. J. Willer, M. Boehnke, K. L. Mohlke, C. M. Lindgren, J. S. Beckmann, I. Barroso, K. E. North, E. Ingelsson, J. N. Hirschhorn, R. J. F. Loos, E. K. Speliotes, Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
11. T. W. Winkler, F. Günther, S. Höllner, M. Zimmermann, R. J. Loos, Z. Kutalik, I. M. Heid, A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. *Nat. Commun.* **9**, 1946 (2018).
 12. I. Verkouter, R. De Mutsert, R. A. J. Smit, S. Trompet, F. R. Rosendaal, D. Van Heemst, K. Willems Van Dijk, R. Noordam, The contribution of tissue-grouped BMI-associated gene sets to cardiometabolic-disease risk: A Mendelian randomization study. *Int. J. Epidemiol.* **49**, 1246–1256 (2020).
 13. G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. Iny Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran, D. Lancet, The GeneCards suite: From gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* **54**, 1.30.1–1.30.33 (2016).
 14. M. Safran, N. Rosen, M. Twik, R. BarShir, T. Iny Stein, D. Dahary, S. Fishilevich, D. Lancet, "The GeneCards Suite" in *Practical Guide to Life Science Databases*, I. Abugessaisa, T. Kasukawa, Eds. (Springer Nature, 2022), pp. 27–56.
 15. E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. Güneş, P. Hall, J. Hayhurst, A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland, K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson, L. Hindorf, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, L. W. Harris, The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).
 16. M. Ghousaini, E. Mountjoy, M. Carmona, G. Peat, E. M. Schmidt, A. Hercules, L. Fumis, A. Miranda, D. Carvalho-Silva, A. Buniello, T. Burdett, J. Hayhurst, J. Baker, J. Ferrer, A. Gonzalez-Uriarte, S. Jupp, M. A. Karim, G. Koscielny, S. MacLlitt-Northen, C. Malangone, Z. M. Pendlington, P. Roncaglia, D. Suveges, D. Wright, O. Vrousou, E. Papa, H. Parkinson, J. A. L. MacArthur, J. A. Todd, J. C. Barrett, J. Schwartzentruber, D. G. Hulcoop, D. Ochoa, E. M. McDonagh, I. Dunham, Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
 17. E. Mountjoy, E. M. Schmidt, M. Carmona, J. Schwartzentruber, G. Peat, A. Miranda, L. Fumis, J. Hayhurst, A. Buniello, M. A. Karim, D. Wright, A. Hercules, E. Papa, E. B. Fauman, J. C. Barrett, J. A. Todd, D. Ochoa, I. Dunham, M. Ghousaini, An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
 18. J. R. Staley, J. Blackshaw, M. A. Kamat, S. Ellis, P. Surendran, B. B. Sun, D. S. Paul, D. Freitag, S. Burgess, J. Danesh, R. Young, A. S. Butterworth, PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
 19. M. A. Kamat, J. A. Blackshaw, R. Young, P. Surendran, S. Burgess, J. Danesh, A. S. Butterworth, J. R. Staley, PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
 20. S. Christakoudi, E. Evangelou, E. Riboli, K. K. Tsilidis, GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer. *Sci. Rep.* **11**, 10688 (2021).
 21. A. Breschi, M. Muñoz-Aguirre, V. Wucher, C. A. Davis, D. Garrido-Martin, S. Djebali, J. Gillis, D. D. Pervouchine, A. Vlasova, A. Dobin, C. Zaleski, J. Drenkow, C. Danyko, A. Scavelli, F. Reverter, M. S. Snyder, T. R. Gingeras, R. Guigó, A limited set of transcriptional programs define major cell types. *Genome Res.* **30**, 1047–1059 (2020).
 22. J. R. Huyghe, S. A. Bien, T. A. Harrison, H. M. Kang, S. Chen, S. L. Schmit, D. V. Conti, C. Qu, J. Jeon, C. K. Edlund, P. Greenside, M. Wainberg, F. R. Schumacher, J. D. Smith, D. M. Levine, S. C. Nelson, N. A. Sinnott-Armstrong, D. Albanes, M. H. Alonso, K. Anderson, C. Arnau-Collell, V. Arndt, C. Bamia, B. L. Banbury, J. A. Baron, S. I. Berndt, S. Bézieau, D. T. Bishop, J. Boehm, H. Boeing, H. Brenner, S. Brezina, S. Buch, D. D. Buchanan, A. Burnett-Hartman, K. Butterbach, B. J. Caan, P. T. Campbell, C. S. Carlson, S. Castellví-Bel, A. T. Chan, J. Chang-Claude, S. J. Chanock, M. D. Chirlaque, S. H. Cho, C. M. Connolly, A. J. Cross, K. Cuk, K. R. Curtis, A. de la Chapelle, K. F. Doheny, D. Duggan, D. F. Easton, S. G. Elias, F. Elliott, D. R. English, E. J. M. Feskens, J. C. Figueiredo, R. Fischer, L. M. FitzGerald, D. Forman, M. Gala, S. Gallinger, W. J. Gauderman, G. G. Giles, E. Gillanders, J. Gong, P. J. Goodman, W. M. Grady, J. S. Grove, A. Gsur, M. J. Gunter, R. W. Haile, J. Hampe, H. Hampel, S. Harlid, R. B. Hayes, P. Hofer, M. Hoffmeister, J. L. Hopper, W. L. Hsu, W. Y. Huang, T. J. Hudson, D. J. Hunter, G. Ibañez-Sanz, G. E. Idos, R. Ingersoll, R. D. Jackson, E. J. Jacobs, M. A. Jenkins, A. D. Joshi, C. E. Joshi, T. O. Keku, T. J. Key, H. R. Kim, E. Kobayashi, L. N. Kolonel, C. Kooperberg, T. Kühn, S. Kúry, S. S. Kweon, S. C. Larsson, C. A. Laurie, L. Le Marchand, S. M. Leal, S. C. Lee, F. Lejbkovicz, M. Lemire, C. I. Li, L. Li, W. Lieb, Y. Lin, A. Lindblom, N. M. Lindor, H. Ling, T. L. Louie, S. Männistö, S. D. Markowitz, V. Martín, G. Masala, C. E. McNeil, M. Melas, R. L. Milne, L. Moreno, N. Murphy, R. Myte, A. Naccarati, P. A. Newcomb, K. Offit, S. Ogino, N. C. Onland-Moret, B. Pardini, P. S. Parfrey, R. Pearlman, V. Perduca, P. D. P. Pharoah, M. Pinchev, E. A. Platz, R. L. Prentice, E. Pugh, L. Raskin, G. Rennert, H. S. Rennert, E. Riboli, M. Rodríguez-Barranco, J. Romm, L. C. Sakoda, C. Schafmayer, R. E. Schoen, D. Seminara, M. Shah, T. Shelford, M. H. Shin, K. Shulman, S. Sierli, M. L. Slattery, M. C. Southey, Z. K. Stadler, C. Stegmaier, Y. R. Su, C. M. Tangen, S. N. Thibodeau, D. C. Thomas, S. S. Thomas, A. E. Toland, A. Trichopoulos, C. M. Ulrich, D. J. Van Den Berg, F. J. B. van Duijnhoven, B. Van Guelpen, H. van Kranen, J. Vijai, K. Viswanathan, P. Vodicka, L. Vodickova, V. Vymetalkova, K. Weigl, S. J. Weinstein, E. White, A. K. Win, C. R. Wolf, A. Wolk, M. O. Woods, A. H. Wu, S. H. Zaidi, B. W. Zanke, Q. Zhang, W. Zheng, P. C. Scacheri, J. D. Potter, M. C. Bassik, A. Kundaje, G. Casey, V. Moreno, G. R. Abecasis, D. A. Nickerson, S. B. Gruber, L. Hsu, U. Peters, Discovery of common and rare genetic risk variants for colorectal cancer. *Nat. Genet.* **51**, 76–87 (2019).
 23. M. Arnold, M. Leitzmann, H. Freisling, F. Bray, I. Romieu, A. Renehan, I. Soerjomataram, Obesity and cancer: An update of the global impact. *Cancer Epidemiol.* **41**, 8–15 (2016).
 24. F. Hu, *Obesity Epidemiology* (Oxford Univ. Press, 2008).
 25. F. Q. Nuttall, Body mass index: Obesity, BMI, and health: A critical review. *Nutr. Today* **50**, 117–128 (2015).
 26. C. J. Bull, J. A. Bell, N. Murphy, E. Sanderson, G. D. Smith, N. J. Timpson, B. L. Banbury, D. Albanes, S. I. Berndt, S. Bézieau, D. T. Bishop, H. Brenner, D. D. Buchanan, A. Burnett-Hartman, G. Casey, S. Castellví-Bel, A. T. Chan, J. Chang-Claude, A. J. Cross, A. de la Chapelle, J. C. Figueiredo, S. J. Gallinger, S. M. Gapstur, G. G. Giles, S. B. Gruber, A. Gsur, J. Hampe, H. Hampel, T. A. Harrison, M. Hoffmeister, L. Hsu, W.-Y. Huang, J. R. Huyghe, M. A. Jenkins, C. E. Joshi, T. O. Keku, T. Kühn, S.-S. Kweon, L. L. Marchand, C. I. Li, L. Li, A. Lindblom, V. Martín, A. M. May, R. L. Milne, V. Moreno, P. A. Newcomb, K. Offit, S. Ogino, A. I. Phipps, E. A. Platz, J. D. Potter, C. Qu, J. R. Quirós, G. Rennert, E. Riboli, L. C. Sakoda, C. Schafmayer, R. E. Schoen, M. L. Slattery, C. M. Tangen, K. K. Tsilidis, C. M. Ulrich, F. J. B. van Duijnhoven, B. van Guelpen, K. Viswanathan, P. Vodicka, L. Vodickova, H. Wang, E. White, A. Wolk, M. O. Woods, A. H. Wu, P. T. Campbell, W. Zheng, U. Peters, E. E. Vincent, M. J. Gunter, Adiposity, metabolites, and colorectal cancer risk: Mendelian randomization study. *BMC Med.* **18**, 396 (2020).
 27. F. Y. Lai, M. Nath, S. E. Hamby, J. R. Thompson, C. P. Nelson, N. J. Samani, Adult height and risk of 50 diseases: A combined epidemiological and genetic analysis. *BMC Med.* **16**, 187 (2018).
 28. N. K. Khankari, X.-O. Shu, W. Wen, P. Kraft, S. Lindström, U. Peters, J. Schildkraut, F. Schumacher, P. Boffetta, A. Risch, H. Bickeböller, C. I. Amos, D. Easton, R. A. Eeles, S. B. Gruber, C. A. Haiman, D. J. Hunter, S. J. Chanock, B. L. Pierce, W. Zheng, Colorectal Transdisciplinary Study (CORECT); Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE); Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE); Transdisciplinary Research in Cancer of the Lung (TRICL), Association between adult height and risk of colorectal, lung, and prostate cancer: Results from meta-analyses of prospective studies and mendelian randomization analyses. *PLOS Med.* **13**, e1002118 (2016).
 29. L. Nunney, Size matters: Height, cell number and a person's risk of cancer. *Proc. Biol. Sci.* **285**, 20181743 (2018).
 30. E. Giovannucci, Nutrition, insulin, insulin-like growth factors and cancer. *Horm. Metab. Res.* **35**, 694–704 (2003).
 31. N. Seyed Khoei, K. H. Wagner, A. M. Sedlmeier, M. J. Gunter, N. Murphy, H. Freisling, Bilirubin as an indicator of cardiometabolic health: A cross-sectional analysis in the UK Biobank. *Cardiovasc. Diabetol.* **21**, 54 (2022).
 32. N. S. Khoei, K. H. Wagner, R. Carreras-torres, M. J. Gunter, N. Murphy, H. Freisling, Associations between prediagnostic circulating bilirubin levels and risk of gastrointestinal cancers in the UK biobank. *Cancers* **13**, 2749 (2021).
 33. E. Esteve, W. Ricart, J. M. Fernández-Real, Dyslipidemia and inflammation: An evolutionary conserved mechanism. *Clin. Nutr.* **24**, 16–31 (2005).
 34. Y. Tian, K. Wang, J. Li, J. Wang, Z. Wang, Y. Fan, Y. Ye, G. Ji, Y. Li, The association between serum lipids and colorectal neoplasm: A systemic review and meta-analysis. *Public Health Nutr.* **18**, 3355–3370 (2015).
 35. X. Yao, Z. Tian, Dyslipidemia and colorectal cancer risk: A meta-analysis of prospective studies. *Cancer Causes Control* **26**, 257–268 (2015).
 36. H. Rodriguez-Broadbent, P. J. Law, A. Sud, K. Palin, S. Tuupanen, A. Gylfe, U. A. Hänninen, T. Cajuso, T. Tanskanen, J. Kondelin, E. Kaasinen, A. P. Sarin, S. Ripatti, J. G. Eriksson, H. Rissanen, P. Knekt, E. Pukkala, P. Jousilahti, V. Salomas, A. Palotie, L. Renkonen-Sinisalo, A. Lepistö, J. Böhm, J. P. Mecklin, N. A. Al-Tassan, C. Palles, L. Martin, E. Barclay, S. M. Farrington, M. N. Timofeeva, B. F. Meyer, S. M. Wakil, H. Campbell, C. G. Smith, S. Idziaszczyk, T. S. Maughan, R. Kaplan, R. Kerr, D. Kerr, M. N. Passarelli, J. C. Figueiredo,

- D. D. Buchanan, A. K. Win, J. L. Hopper, M. A. Jenkins, N. M. Lindor, P. A. Newcomb, S. Gallinger, D. Conti, F. Schumacher, G. Casey, L. A. Aaltonen, J. P. Chadler, I. P. Tomlinson, M. G. Dunlop, R. S. Houlston, Mendelian randomisation implicates hyperlipidaemia as a risk factor for colorectal cancer. *Int. J. Cancer* **140**, 2701–2708 (2017).
37. G. Ibáñez-Sanz, A. Díez-Villanueva, M. Riera-Ponsati, T. Fernández-Villa, P. Fernández-Navarro, M. Bustamante, J. Llorca, P. Amiano, N. Ascunce, G. Fernández-Tardón, I. Salcedo Bellido, D. Salas, R. Capelo Álvarez, M. Crous-Bou, L. Ortega-Valín, B. Pérez-Gómez, G. Castaño-Vinyals, C. Palazuelos, J. M. Altzibar, E. Ardanaz, A. Tardón, J. J. Jiménez Moleón, V. Olmos Juste, N. Aragonés, M. Pollán, M. Kogevinas, V. Moreno, Mendelian randomization analysis rules out dyslipidaemia as colorectal cancer cause. *Sci. Rep.* **9**, 13407 (2019).
38. A. Gómez-Hernández, N. Beneit, S. Díaz-Castroverde, Ó. Escribano, Differential role of adipose tissues in obesity and related metabolic and vascular complications. *Int. J. Endocrinol.* **2016**, 1216783 (2016).
39. G. J. Farkas, D. R. Gater, Neurogenic obesity and systemic inflammation following spinal cord injury: A review. *J. Spinal Cord Med.* **41**, 378–387 (2018).
40. R. J. F. Loos, T. O. Kilpeläinen, Genes that make you fat, but keep you healthy. *J. Intern. Med.* **284**, 450–463 (2018).
41. B. Brumpton, E. Sanderson, K. Heilbron, F. P. Hartwig, S. Harrison, G. Å. Vie, Y. Cho, L. D. Howe, A. Hughes, D. I. Boomsma, A. Havdahl, J. Hopper, M. Neale, M. G. Nivard, N. L. Pedersen, C. A. Reynolds, E. M. Tucker-Drob, A. Grotzinger, L. Howe, T. Morris, S. Li, A. Auton, F. Windmeijer, W. M. Chen, J. H. Bjørnsgaard, K. Hveem, C. Willer, D. M. Evans, J. Kaprio, G. Davey-Smith, B. O. Åsvold, G. Hemani, N. M. Davies, Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat. Commun.* **11**, 3519 (2020).
42. UK Biobank Coordinating Centre, *UK Biobank: Protocol for a Large-Scale Prospective Epidemiological Resource* (UK Biobank Coordinating Centre, 2007).
43. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
44. N. J. Salkind, Ed., "Winsorize," in *Encyclopedia of Research Design* (Sage Publications Inc., 2012).
45. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. De Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
46. C. C. Chang, C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
47. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
48. RStudio Team, *RStudio: Integrated Development Environment for R* (RStudio, 2022).
49. P.-R. Loh, G. Kichaev, S. Gazal, A. P. Schoech, A. L. Price, Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
50. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang; Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
51. Hemani G. ieuGWAS: R interface to the IEU GWAS database API. R package (2020).
52. B. Elsworth, M. Lyon, T. Alexander, Y. Liu, P. Matthews, J. Hallett, P. Bates, T. Palmer, W. Haberland, G. Davey, S. 1, J. Zheng, P. Haycock, T. R. Gaunt, G. Hemani, The MRX IEU OpenGWAS data infrastructure. bioRxiv 2020.08.10.244293 [Preprint] (2020); <https://doi.org/10.1101/2020.08.10.244293>.
53. G. Hemani, J. Zheng, B. Elsworth, K. H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J. Bowden, R. Langdon, V. Y. Tan, J. Yarmolinsky, H. A. Shihab, N. J. Timpson, D. M. Evans, C. Relton, R. M. Martin, G. Davey Smith, T. R. Gaunt, P. C. Haycock, The MR-base platform supports systematic causal inference across the human phenotype. *eLife* **7**, e34408 (2018).
54. L. Yengo, S. Vedantam, E. Marouli, J. Sidorenko, E. Bartell, S. Sakaue, M. Graff, A. U. Eliassen, Y. Jiang, S. Raghavan, J. Miao, J. D. Arias, S. E. Graham, R. E. Mukamel, C. N. Spracklen, X. Yin, S.-H. Chen, T. Ferreira, H. H. Highland, Y. Ji, T. Karaderi, K. Lin, K. Lüll, D. E. Malden, C. Medina-Gomez, M. Machado, A. Moore, S. Rieger, X. Sim, S. Vrieze, T. S. Ahluwalia, M. Akiyama, M. A. Allison, M. Alvarez, M. K. Andersen, A. Ani, V. Appadurai, L. Arbeeva, S. Bhaskar, L. F. Bielak, S. Bollepalli, L. L. Bonnyycastle, J. Bork-Jensen, J. P. Bradfield, Y. Bradford, P. S. Braund, J. A. Brody, K. S. Burgdorf, B. E. Cade, H. Cai, Q. Cai, A. Campbell, M. Cañadas-Garre, E. Catamo, J.-F. Chai, X. Chai, L.-C. Chang, Y.-C. Chang, C.-H. Chen, A. Chesí, S. H. Choi, R.-H. Chung, M. Cocca, M. P. Concas, C. Couture, G. Cuellar-Partida, R. Danning, E. W. Daw, F. Degehard, G. E. Delgado, A. Delitala, A. Demirkan, X. Deng, P. Devineni, A. Dietl, M. Dimitriou, L. Dimitrov, R. Dorajoo, A. B. Ekici, J. E. Engmann, Z. Fairhurst-Unter, A.-E. Farmaki, J. D. Faul, J.-C. Fernandez-Lopez, L. Forer, M. Francescotto, S. Freitag-Wolf, C. Fuchsberger, T. E. Galesloot, Y. Gao, Z. Gao, F. Geller, O. Giannakopoulou, F. Giulianini, A. P. Gjesing, A. Goel, S. D. Gordon, M. Gorski, J. Grove, X. Guo, S. Gustafsson, J. Haessler, T. F. Hansen, A. S. Havulinna, S. J. Haworth, J. He,
- N. Heard-Costa, P. Hebbard, G. Hindy, Y.-L. A. Ho, E. Hofer, E. Holliday, K. Horn, W. E. Hornsby, J.-J. Hottenga, H. Huang, J. Huang, A. Huerta-Chagoya, J. E. Huffman, Y.-J. Hung, S. Huo, M. Y. Hwang, H. Iha, D. D. Ikeda, M. Isono, A. U. Jackson, S. Jäger, I. E. Jansen, I. Johansson, J. B. Jonas, A. Jonsson, T. Jørgensen, I.-P. Kalafati, M. Kanai, S. Kanoni, L. L. Kårhus, A. Kasturiratne, T. Katsuya, T. Kawaguchi, R. L. Kember, K. A. Kentistou, H.-N. Kim, Y. J. Kim, M. E. Kleber, M. J. Knol, A. Kurbasic, M. Lauzon, P. Le, R. Lea, J.-Y. Lee, H. L. Leonard, S. A. Li, X. Li, X. Li, J. Liang, H. Lin, S.-Y. Lin, J. Liu, X. Liu, K. S. Lo, J. Long, L. Lores-Motta, J. Luan, V. Lyssenko, L.-P. Lytykäinen, A. Mahajan, V. Mamakou, M. Mangino, A. Manichaikul, J. Marten, M. Mattheisen, L. Mavarani, A. F. McDaid, K. Meidtnet, R. L. Melendez, J. M. Mercader, Y. Milanecchi, J. E. Miller, I. Y. Millwood, P. P. Mishra, R. E. Mitchell, L. M. Tellechea, A. Morgan, S. Mucha, M. Munz, M. Nakatочи, C. P. Nelson, M. Nethander, C. W. Nho, A. A. Nielsen, I. M. Nolte, S. S. Nongmaithem, R. Noordam, I. Ntalla, T. Nutile, A. Pandit, P. Christofidou, K. Pärna, M. Pauper, E. R. B. Petersen, L. V. Petersen, R. Pitkännén, O. Polašek, A. Poveda, M. H. Preuss, S. Pyarajan, L. M. Raffield, H. Rakugi, J. Ramirez, A. Rasheed, D. Raven, N. W. Rayner, C. Rivas, R. Rohde, D. Ruggiero, S. E. Ruotsalainen, K. A. Ryan, M. Sabater-Lleal, R. Saxena, M. Scholz, A. Sendamarai, B. Shen, J. Shi, J. H. Shin, C. Sidore, C. M. Sitlani, R. C. Sliker, R. A. J. Smit, A. V. Smith, J. A. Smith, L. J. Smyth, L. Southam, V. Steinthorsdottir, L. Sun, F. Takeuchi, D. S. P. Tallapragada, K. D. Taylor, B. O. Tayo, C. Tcheandjieu, N. Terzikhan, P. Tesolin, A. Teumer, E. Theusch, D. J. Thompson, G. Thorleifsson, P. R. H. J. Timmers, S. Trompet, C. Turman, S. Vaccargiu, S. W. van der Laan, P. J. van der Mer, J. B. van Klínen, J. van Setten, S. S. Verma, N. Verweij, Y. Veruri, C. A. Wang, C. Wang, L. Wang, Z. Wang, H. R. Warren, W. B. Wei, A. R. Wickremesinghe, M. Wielscher, K. L. Wiggins, B. S. Winsvold, A. Wong, Y. Wu, M. Wuttke, R. Xia, T. Xie, K. Yamamoto, J. Yang, J. Yao, H. Young, N. A. Younsri, L. Yu, L. Zeng, W. Zhang, X. Zhang, J.-H. Zhao, W. Zhao, W. Zhou, M. E. Zimmermann, M. Zoledziwska, L. S. Adair, H. H. H. Adams, C. A. Aguilar-Salinas, F. Al-Mulla, D. K. Arnett, F. W. Asselbergs, B. O. Åsvold, J. Attia, B. Banas, S. Bandinelli, D. A. Bennett, T. Bergler, D. Bharadwaj, G. Biino, H. Bisgaard, E. Boerwinkle, C. A. Böger, K. Bønnelykke, D. I. Boomsma, A. D. Børghlum, J. B. Borja, C. Bouchard, D. W. Bowden, I. Brandslund, B. Brumpton, J. E. Buring, M. J. Caulfield, J. C. Chambers, G. R. Chandak, S. J. Chanock, N. Chaturvedi, Y.-D. I. Chen, Z. Chen, C.-Y. Cheng, I. E. Christophersen, M. Cui, J. W. Cole, F. S. Collins, R. S. Cooper, M. Cruz, F. Cucca, L. A. Cupples, M. J. Cutler, S. M. Damrauer, T. M. Dantoft, G. J. de Borst, L. C. P. G. de Groot, P. L. De Jager, D. P. V. de Kleijn, H. J. de Silva, G. V. Dedoussis, A. I. den Hollander, S. Du, D. F. Easton, P. J. M. Elders, A. H. Eliassen, P. T. Ellinor, S. Elmståhl, J. Erdmann, M. K. Evans, D. Fatkin, B. Feenstra, M. F. Feitosa, L. Ferrucci, I. Ford, M. Fornage, A. Franke, P. W. Franks, B. I. Freedman, P. Gasparini, C. Gieger, G. Grotto, M. E. Goddard, Y. M. Goltightly, C. Gonzalez-Villalpando, P. Gordon-Larsen, H. Grallert, S. F. A. Grant, N. Grarup, L. Griffiths, V. Gudnason, C. Haiman, H. Hakonarson, T. Hansen, C. A. Hartman, A. T. Hattersley, C. Hayward, S. R. Heckbert, C.-K. Heng, C. Hengstenberg, A. W. Hewitt, H. Hishigaki, C. B. Hoyng, P. L. Huang, W. Huang, S. C. Hunt, K. Hveem, E. Hyppönen, W. G. Iacono, S. Ichihara, M. A. Ikram, C. R. Isasi, R. D. Jackson, M.-R. Jarvelin, Z.-B. Jin, K.-H. Jöckel, P. K. Joshi, P. Jousilahti, J. W. Jukema, C. M. Kähönen, Y. Kamatani, K. D. Kang, J. Kaprio, S. L. R. Kardia, F. Karpe, N. Kato, F. Kee, T. Kessler, A. V. Khera, C. C. Khor, L. A. L. M. Kiemeny, B.-J. Kim, E. K. Kim, H.-L. Kim, P. Kirchhof, M. Kivimaki, W.-P. Koh, H. A. Koistinen, G. D. Kolovou, J. S. Kooper, C. Kooperberg, A. Köttgen, P. Kovacs, A. Kraaijeveld, P. Kraft, R. M. Krauss, M. Kumari, Z. Kutalik, M. Laakso, L. A. Lange, C. Langenberg, L. J. Launer, L. L. Marchand, H. Lee, N. R. Lee, T. Lehtimäki, H. Li, L. Li, W. Lieb, X. Lin, L. Lind, A. Linneberg, C.-T. Liu, J. Liu, M. Loeffler, B. London, S. A. Lubitz, S. J. Lye, D. A. Mackey, R. Mägi, P. K. E. Magnusson, G. M. Marcus, P. M. Vidal, N. G. Martin, W. März, F. Matsuda, R. W. McFarrah, M. M. Gue, A. J. McKnight, S. E. Medland, D. Mellström, A. Metspalu, B. D. Mitchell, P. Mitchell, D. O. Mook-Kanamori, A. D. Morris, L. A. Mucci, P. B. Munroe, M. A. Nalls, S. Nazarian, A. E. Nelson, M. J. Neville, C. Newton-Cheh, C. S. Nielsen, M. M. Nöthen, C. Ohlsson, A. J. Oldehinkel, L. Orozco, K. Pakkala, P. Pajukanta, C. N. A. Palmer, E. J. Parra, C. Pattaro, O. Pedersen, C. E. Pennell, B. W. J. H. Penninx, L. Perusse, A. Peters, P. A. Peyser, D. J. Porteous, D. Posthuma, C. Power, P. P. Pramstaller, M. A. Province, Q. Qi, J. Qu, D. J. Rader, O. T. Raitakari, S. Ralhan, L. S. Rallidis, D. C. Rao, S. Redline, D. F. Reilly, A. P. Reiner, S. Y. Rhee, P. M. Ridker, M. Rienstra, S. Ripatti, M. D. Ritchie, D. M. Roden, F. R. Rosendaal, J. I. Rotter, I. Rudan, F. Rutter, C. Sabanayagam, D. Saleheen, V. Salomaa, N. J. Samani, D. K. Sanghera, N. Sattar, B. Schmidt, H. Schmidt, R. Schmidt, M. B. Schulze, H. Schunkert, L. J. Scott, R. J. Scott, P. Sever, E. J. Shiroma, M. B. Shoemaker, X.-O. Shu, E. M. Simonsick, M. Sims, J. R. Singh, A. B. Singleton, M. F. Sinner, J. G. Smith, H. Snieder, T. D. Spector, M. J. Stampfer, K. J. Stark, D. P. Strachan, L. M. 't Hart, Y. Tabara, H. Tang, J. C. Tardif, T. A. Thanaraj, N. J. Timpson, A. Tönjes, A. Tremblay, T. Tuomi, J. Tuomilehto, M.-T. Tusié-Luna, A. G. Uitterlinden, R. M. van Dam, P. van der Harst, N. Van der Velde, C. M. van Duijn, N. M. van Schoor, V. Vitart, U. Völker, P. Vollenweider, H. Völzke, N. H. Wachter-Rodarte, M. Walker, Y. X. Wang, N. J. Wareham, R. M. Watanabe, H. Watkins, D. R. Weir, T. M. Werge, E. Widen, L. R. Wilkens, G. Willemssen, W. C. Willett, J. F. Wilson, T.-Y. Wong, J.-T. Woo, A. F. Wright, J.-Y. Wu, H. Xu, C. S. Yajnik, M. Yokota, J.-M. Yuan, E. Zeggini, B. S. Zemel, W. Zheng, X. Zhu, J. M. Zmuda, A. B. Zonderman, J.-A. Zwart; 23andMe Research Team; VA Million Veteran Program; DiscovEHR (DiscoveHR, MyCommunity Health Initiative); eMERGE (Electronic Medical Records, Genomics Network); Lifelines Cohort Study; The

- PRACTICAL Consortium; Understanding Society Scientific Group, D. I. Chasman, Y. S. Cho, I. M. Heid, M. I. McCarthy, M. C. Y. Ng, C. J. O'Donnell, F. Rivadeneira, U. Thorsteinsdottir, Y. V. Sun, E. S. Tai, M. Boehnke, P. Deloukas, A. E. Justice, C. M. Lindgren, R. J. F. Loos, K. L. Mohlke, K. E. North, K. Stefansson, R. G. Walters, T. W. Winkler, K. L. Young, P.-R. Loh, J. Yang, T. Esko, T. L. Assimes, A. Auton, G. R. Abecasis, C. J. Willer, A. E. Locke, S. I. Berndt, G. Lettre, T. M. Frayling, Y. Okada, A. R. Wood, P. M. Visscher, J. N. Hirschhorn, A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).
55. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: Generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
56. K. Watanabe, E. Taskesen, A. Van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
57. GTEx Consortium, The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
58. S. Burgess, F. Dudbridge, S. G. Thompson, Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Stat. Med.* **35**, 1880–1906 (2016).
59. S. Burgess, S. G. Thompson, Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
60. G. D. Smith, S. Ebrahim, "Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
61. D. A. Lawlor, R. M. Harbord, J. A. C. Sterne, N. Timpson, G. D. Smith, Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
62. V. Didelez, N. Sheehan, Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Methods Med. Res.* **16**, 309–330 (2007).
63. F. P. Hartwig, N. M. Davies, G. Hemani, G. Davey Smith, Two-sample Mendelian randomization: Avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2016).
64. J. Bowden, F. D. G. M. C. Minelli, G. D. Smith, N. Sheehan, J. Thompson, A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* **36**, 1783–1802 (2017).
65. N. L. Dimou, K. K. Tsilidis, A primer in Mendelian randomization methodology with a focus on utilizing published summary association data. *Methods Mol. Biol.* **1793**, 211–230 (2018).
66. J. P. T. Higgins, S. G. Thompson, J. J. Deeks, D. G. Altman, Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
67. J. Bowden, G. D. Smith, S. Burgess, Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
68. P. Pagoni, N. L. Dimou, N. Murphy, E. Stergiakouli, Using Mendelian randomisation to assess causality in observational studies. *Evid. Based Ment. Health* **22**, 67–71 (2019).
69. J. Bowden, G. Davey Smith, P. C. Haycock, S. Burgess, Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
70. O. O. Yavorska, S. Burgess, MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
71. World Health Organization, *International Classification of Diseases for Oncology (ICD-O): First Revision* (World Health Organization, ed. 3, 2013).
72. M. A. Hernán, S. Hernández-Díaz, J. M. Robins, A structural approach to selection bias. *Epidemiology* **15**, 615–625 (2004).
73. S. Greenland, J. Pearl, J. M. Robins, Causal diagrams for epidemiologic research. *Epidemiology* **10**, 37–48 (1999).
74. P. Hepsomali, J. A. Groeger, Diet and general cognitive ability in the UK Biobank dataset. *Sci. Rep.* **11**, 11786 (2021).
75. D. F. Moore, *Applied Survival Analysis Using R* (Springer, 2016).
76. S. Lê, J. Josse, F. Husson, FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
77. F. E. Harrel Jr., *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (Springer, 2015).

Acknowledgments: UKB is an open access resource. Bona fide researchers can apply to use the UKB dataset by registering and applying at <http://ukbiobank.ac.uk/register-apply/>. This research has been conducted using the UKB Resource under application number 55870, and we express our gratitude to the participants and those involved in building the resource. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Additional acknowledgements are given in Supplementary Text. **Funding:** This work was supported by the French National Cancer Institute (INCA_14108) and the German Research Foundation (BA 5459/2-1). The National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC) provided infrastructure support for the Department of Epidemiology and Biostatistics at Imperial College London (UK). Study-specific funding sources are given in Supplementary Text. **Author contributions:** Conceptualization: L.P.-N. and H.F. Methodology: L.P.-N. and A.M.S. Investigation: L.P.-N. and A.M.S. Visualization: L.P.-N. Supervision: R.C.-T., M.J.G., and H.F. Writing—original draft: L.P.-N. and A.M.S. Writing—review and editing: L.P.-N., A.M.S., N.D., H.Ba., B.F., E.F., J.K., K.K.T., S.C., A.J., R.C., P.B., M.J.S., A.W., S.B., H.Br., A.T.C., I.C., J.C.F., K.G.-E., V.M., C.C.N., S.L.S., M.S., C.M.U., P.F., V.V., R.C.-T., M.J.G., and H.F. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The UK Biobank resource is available to bona fide researchers for health-related research in the public interest. All researchers who wish to access the research resource must register with UK Biobank by completing the registration form in the Access Management System (AMS- <https://ams.ukbiobank.ac.uk/ams/>). The UK Biobank data can be provided by UK Biobank Limited pending scientific review and a completed material transfer agreement. The transcriptome data information of human subjects can be downloaded from the GTEx portal (<https://gtexportal.org/home/>, GTEx v8 release phs000424.v8.p8). Genotype data from CCFR, CORECT, and GECCO are available through dbGaP (accession nos. phs001415.v1.p1, phs001315.v1.p1, phs001078.v1.p1, phs001903.v1.p1, phs001856.v1.p1, and phs001045.v1.p1). Requests for the genotype data stratified by colorectal cancer subtypes and by men and women should be submitted to gecco@fredhutch.org. The codes for the computation of the body shape phenotypes, the observational and Mendelian randomization analyses, the GWAS, and the related documentation are available at <https://doi.org/10.5281/zenodo.10063272> and at GitHub (https://github.com/LaiaPeruchet/Body_Shape_Colorectal_Cancer).

Submitted 13 June 2023

Accepted 12 March 2024

Published 19 April 2024

10.1126/sciadv.adj1987