Science Advances
AAAS

Supplementary Materials for

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

Laia Peruchet-Noray *et al.*

Corresponding author: Heinz Freisling, freislingh@iarc.who.fr; Marc J. Gunter, m.gunter@imperial.ac.uk; Robert Carreras-Torres, rcarreras@idibgi.org

Sci. Adv. **10**, eadj1987 (2024)
DOI: 10.1126/sciadv.adj1987

The PDF file includes:

Figs. S1 to S17
Tables S1 to S10
Legend for data S1
Supplementary Text

Other Supplementary Material for this manuscript includes the following:

Data S1

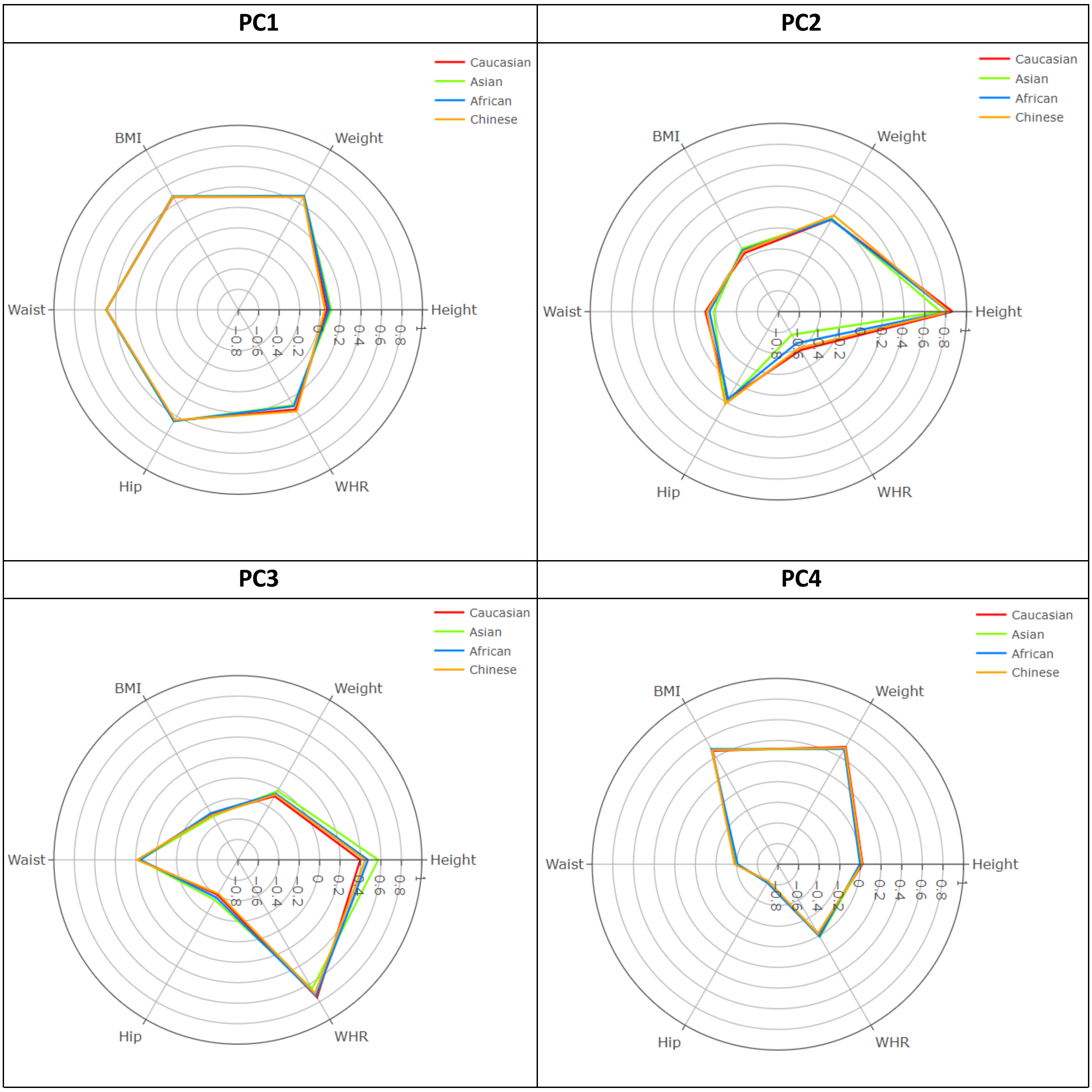

Figure S1. Body shape phenotypes (PC1 - 4) loadings by different ancestries available in UK Biobank (Caucasian in red, Asian in green, African in blue, and Chinese in orange). $N_{Caucasians}$ = *460,198*, N_{African} = 7,269, N_{Asian} = 9,751 and N_{Chinese} = 2,359. PC, principal component.

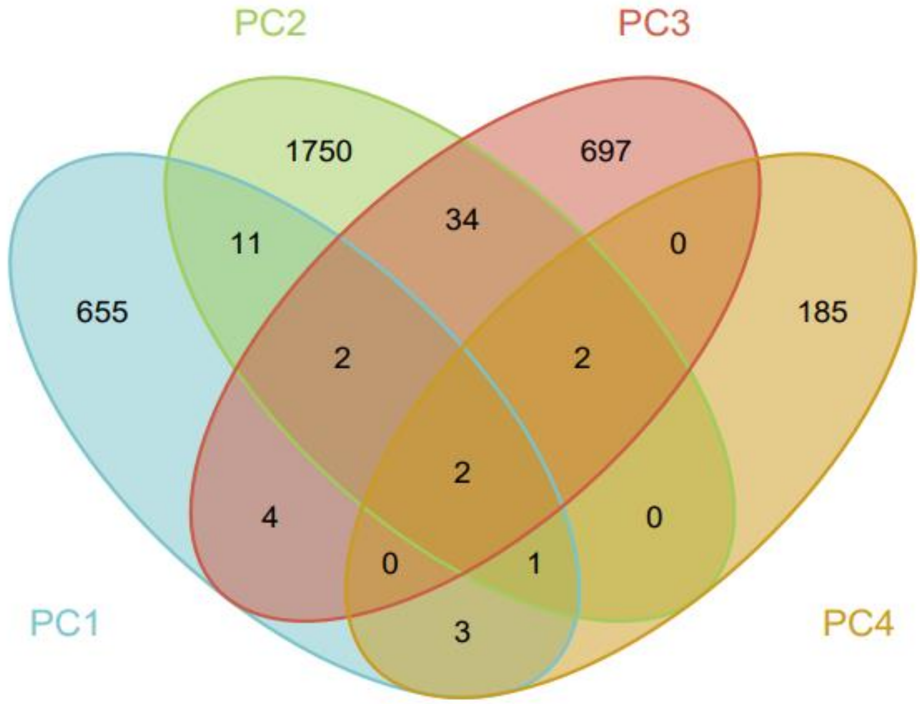

Figure S2. Venn diagram showing the number of genetic variants associated to each body shape phenotype (PC1 - 4) and the genetic variants they share. PC, principal component.

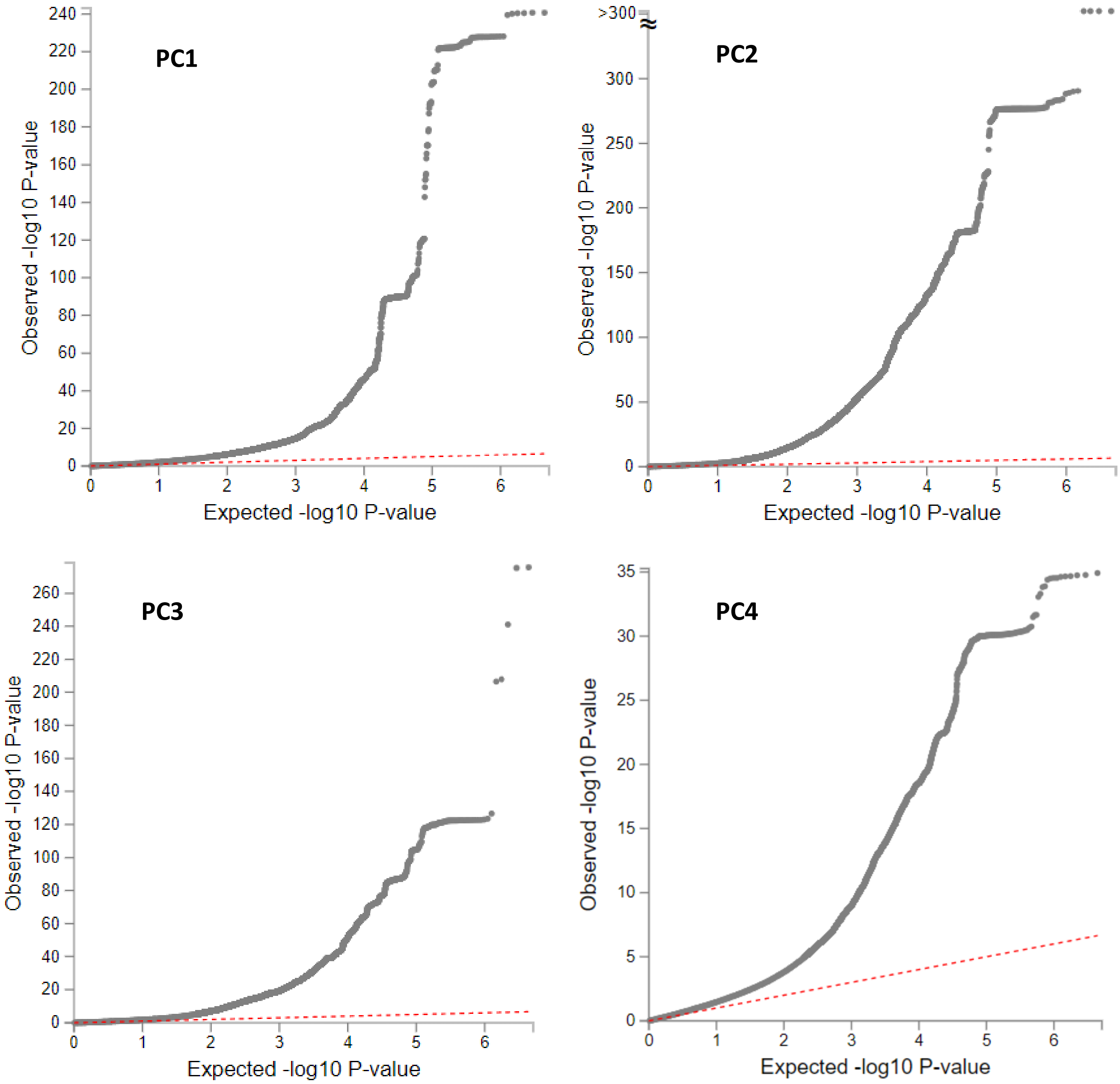

Figure S3. Q-Q plots of body shape phenotypes (PC1 - 4) genome-wide association study (GWAS) summary statistics. PC, principal component.

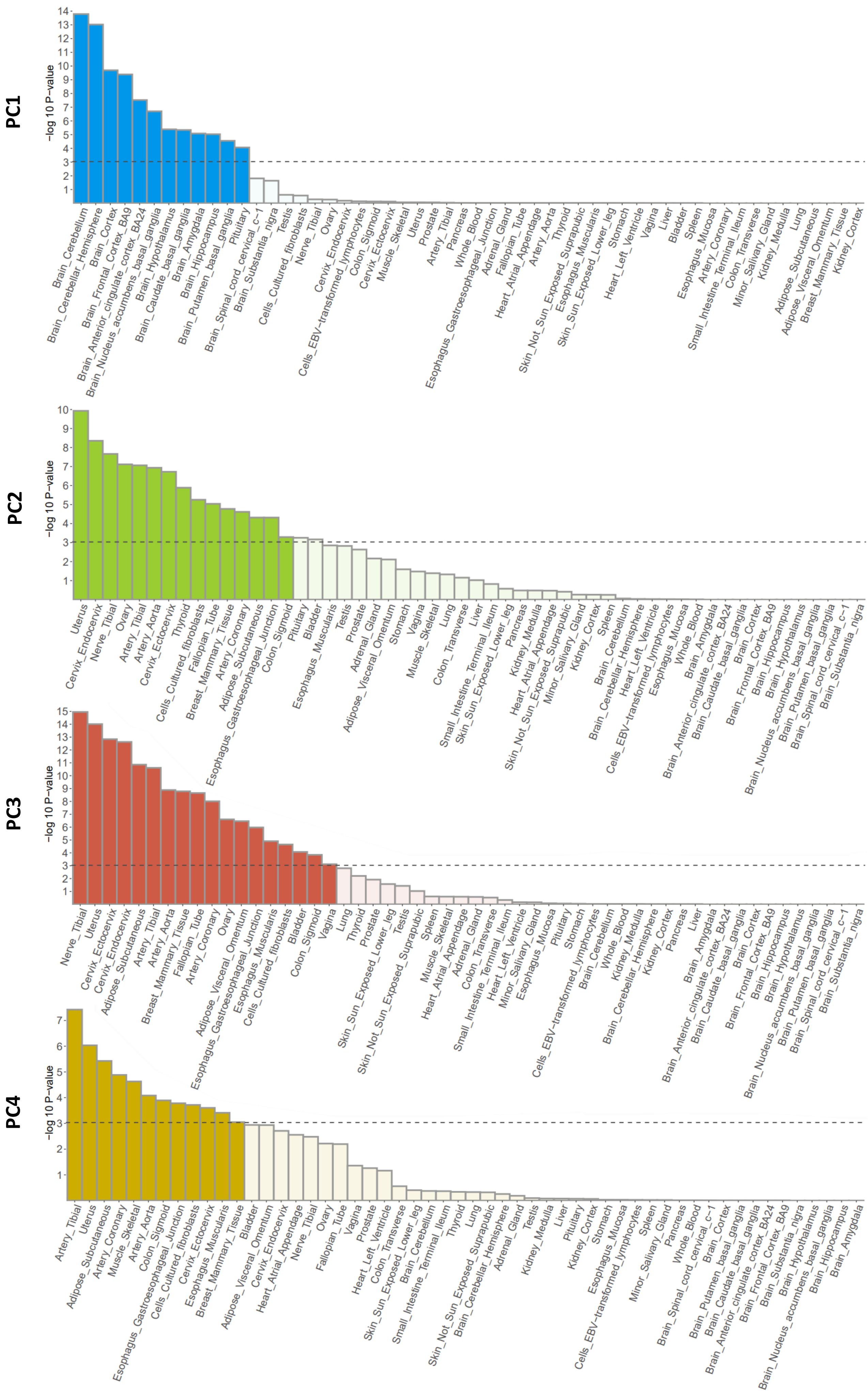

Figure S4. Specific tissue expression profile analysis for the four body shape phenotypes (PC1 - 4). Specific tissues reaching a p-value < 9.26×10^{-4}, after Bonferroni correction, are highlighted in darker colors. PC, principal component.

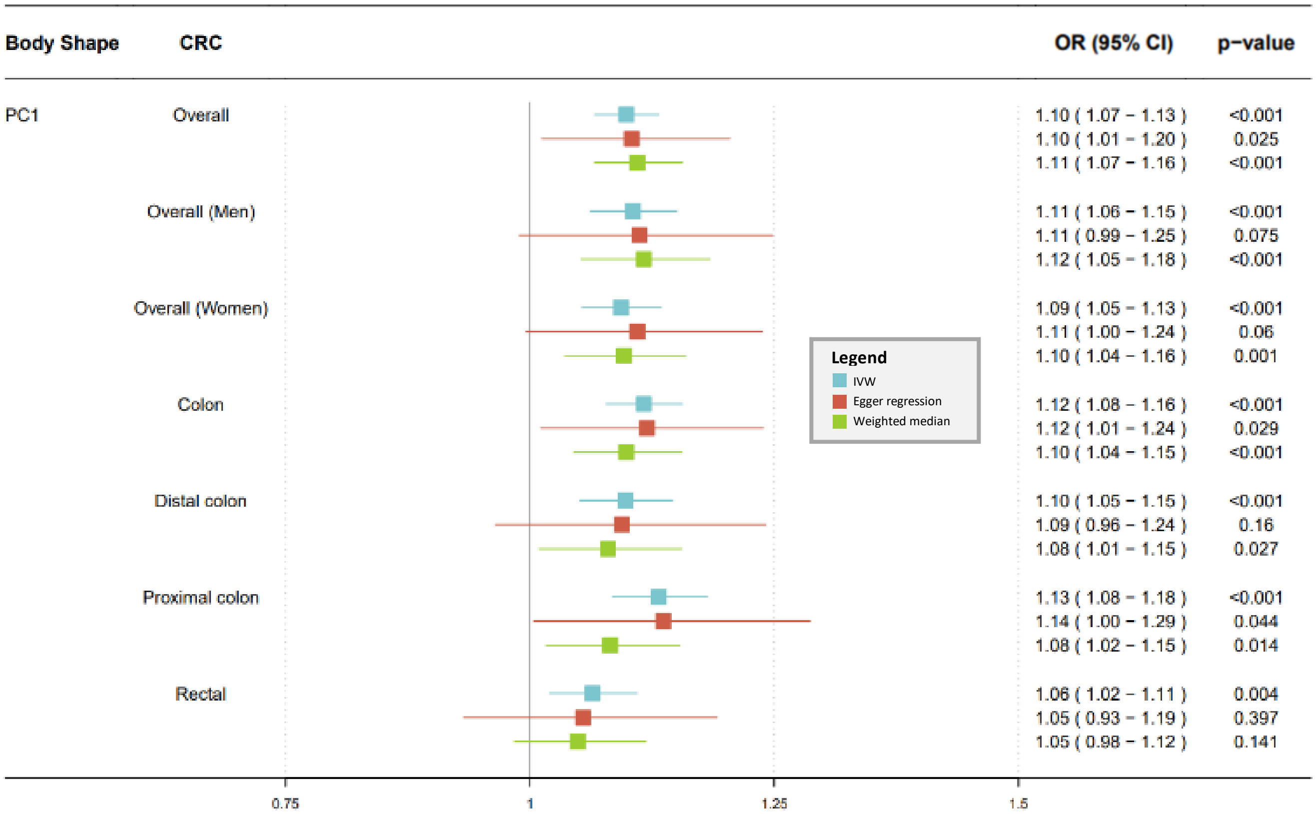

Figure S5. Mendelian randomization (MR) estimates between body shape phenotype PC1 and colorectal cancer (CRC) risk, by sex, and subsites using IVW (blue), MR Egger regression (red) and weighted median (green) MR approaches. CI, confidence interval; OR, odds ratio; PC, principal component.

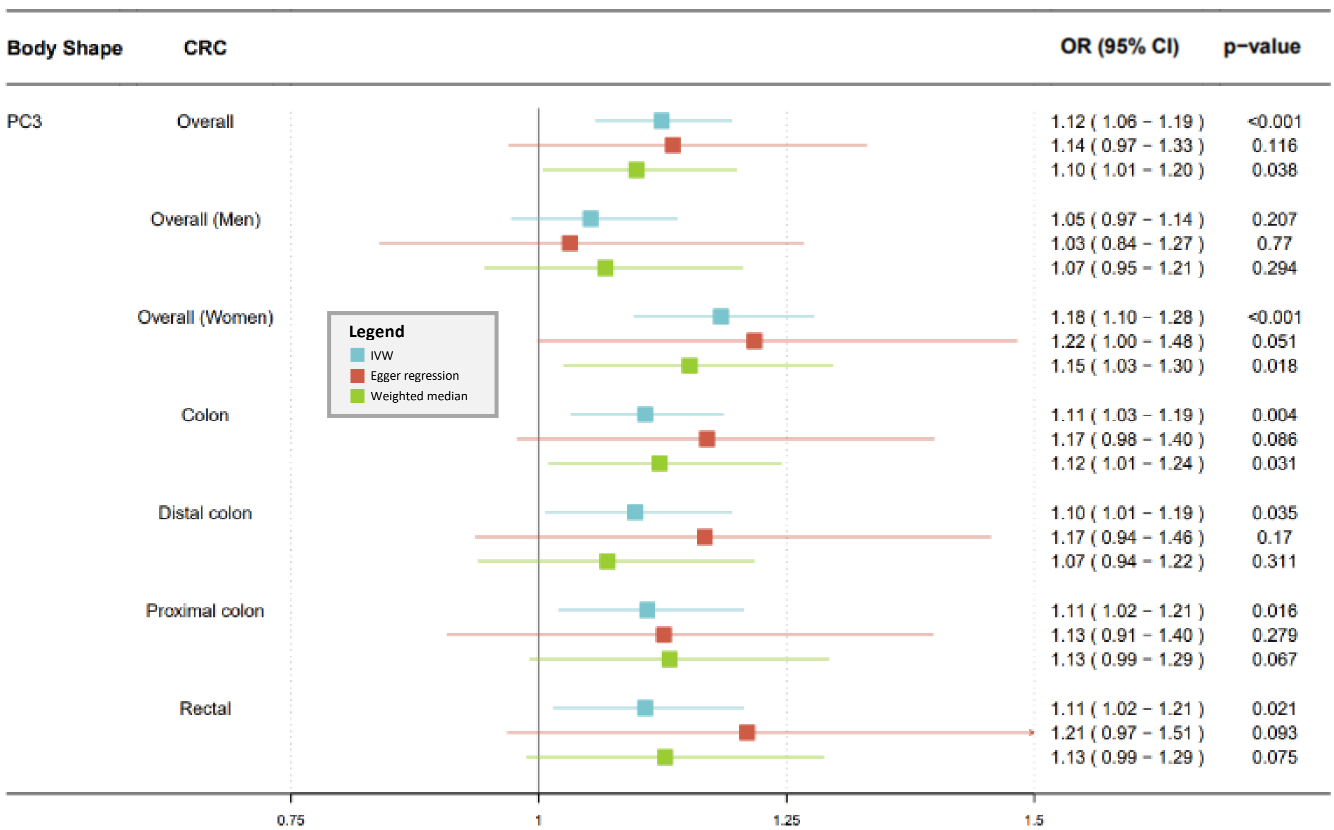

Figure S6. Mendelian randomization (MR) estimates between body shape phenotype PC3 and colorectal cancer (CRC) risk, by sex, and subsites using random-effects inverse-variance weighted (IVW) (blue), MR Egger regression (red) and weighted median (green) MR approaches. CI, confidence interval; OR, odds ratio; PC, principal component.

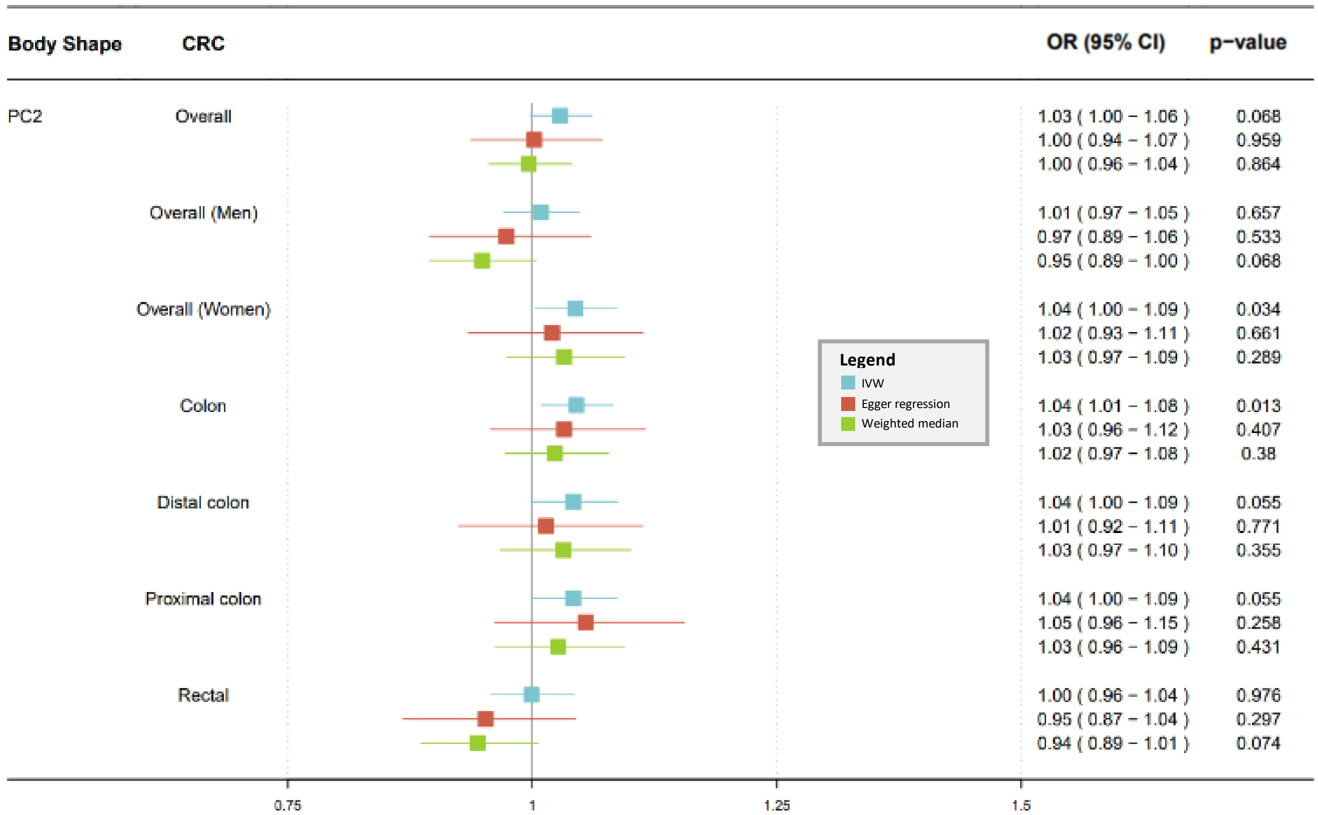

Figure S7. Mendelian randomization (MR) estimates between body shape phenotype PC2 and colorectal cancer (CRC) risk, by sex, and random-effects inverse-variance weighted (IVW) (blue), MR Egger regression (red) and weighted median (green) MR approaches. CI, confidence interval; OR, odds ratio; PC, principal component.

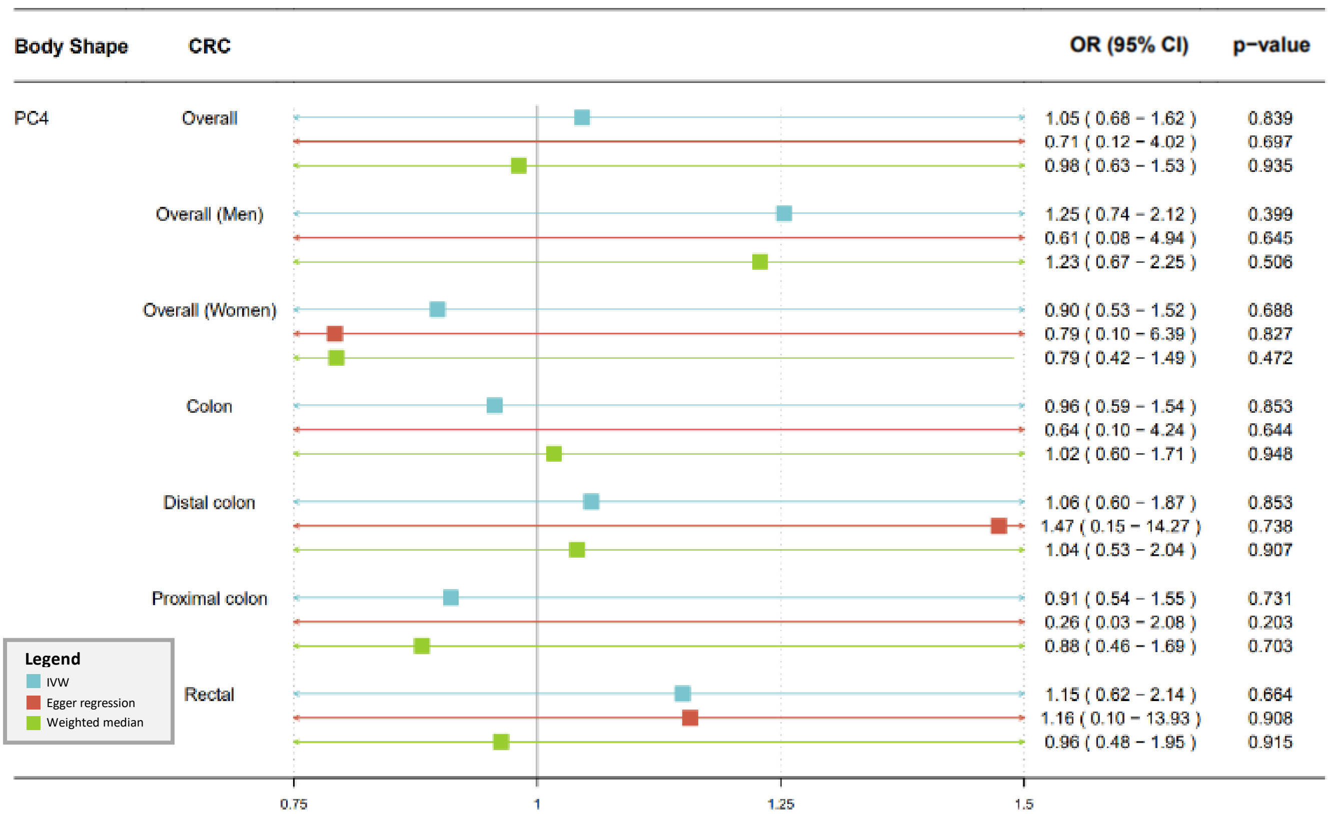

Figure S8. Mendelian randomization (MR) estimates between body shape phenotype PC4 and colorectal cancer (CRC) risk, by sex, and subsites random-effects inverse-variance weighted (IVW) (blue), MR Egger regression (red) and weighted median (green) MR approaches. CI, confidence interval; OR, odds ratio; PC, principal component.

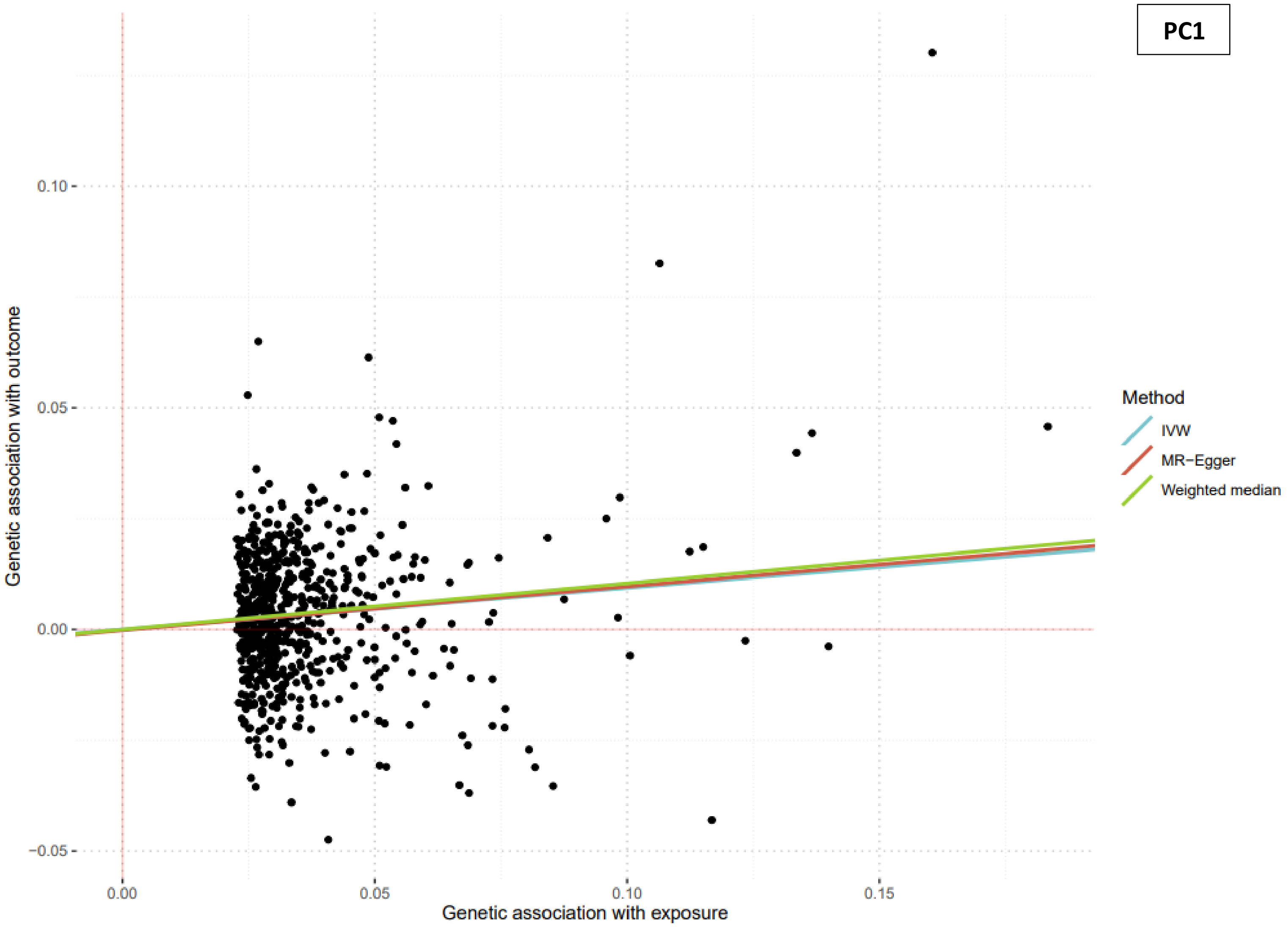

Figure S9. Scatter plot depicting the genetic associations of the body shape PC1 related instruments with the risk of overall colorectal cancer. PC, principal component.

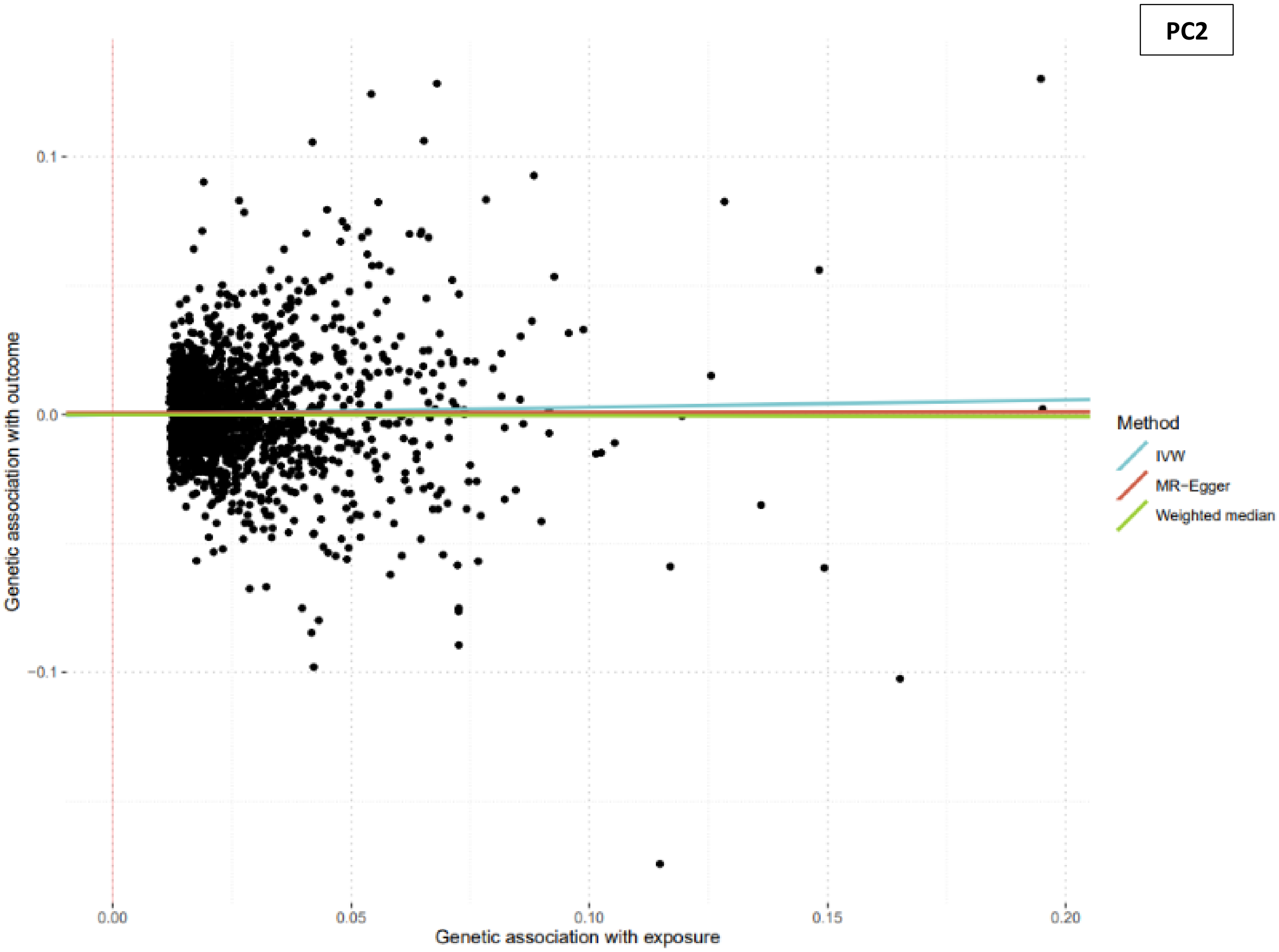

Figure S10. Scatter plot depicting the genetic associations of the body shape PC2 related instruments with the risk of overall colorectal cancer. PC, principal component.

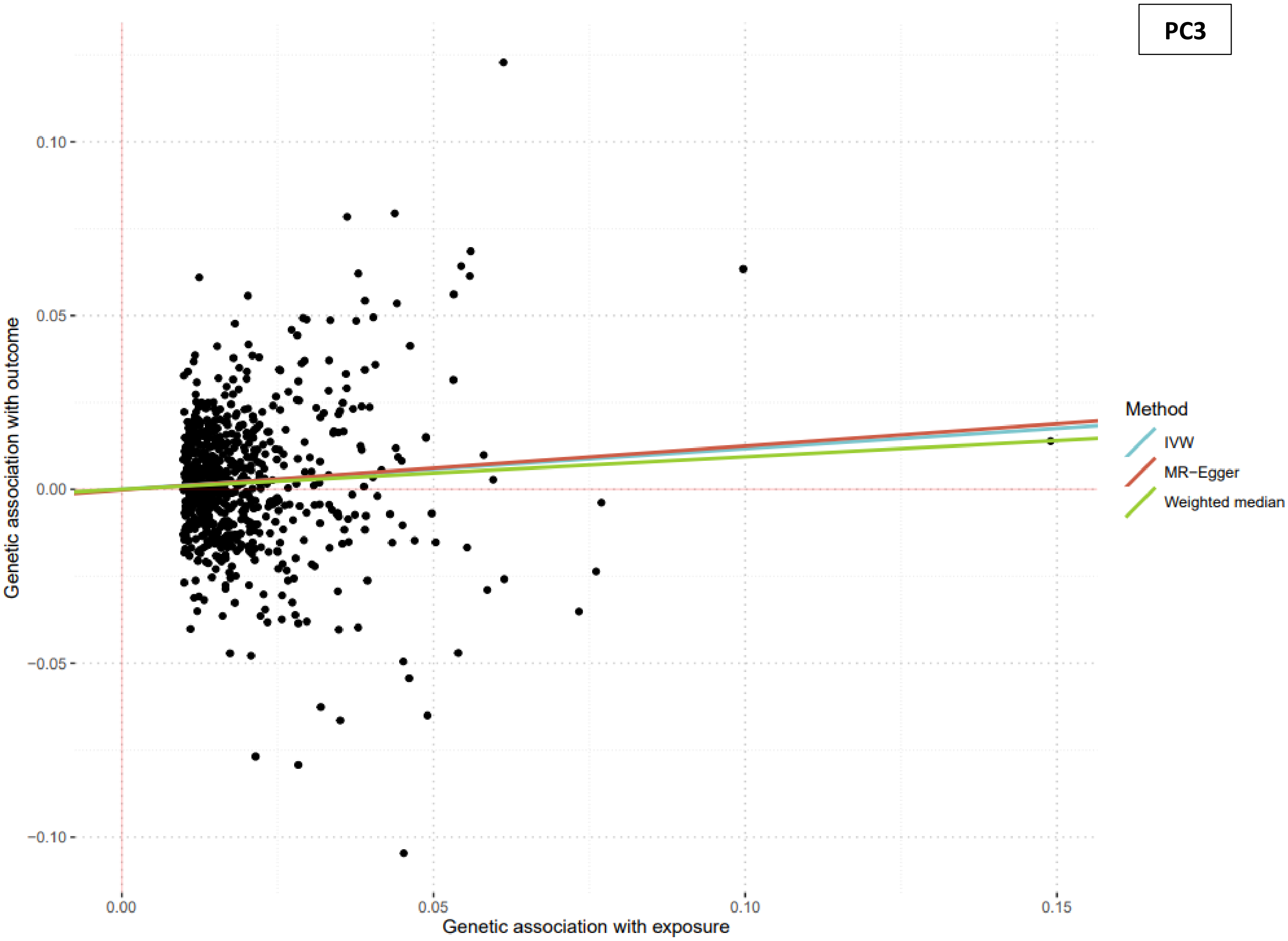

Figure S11. Scatter plot depicting the genetic associations of the body shape PC3 related instruments with the risk of overall colorectal cancer. PC, principal component.

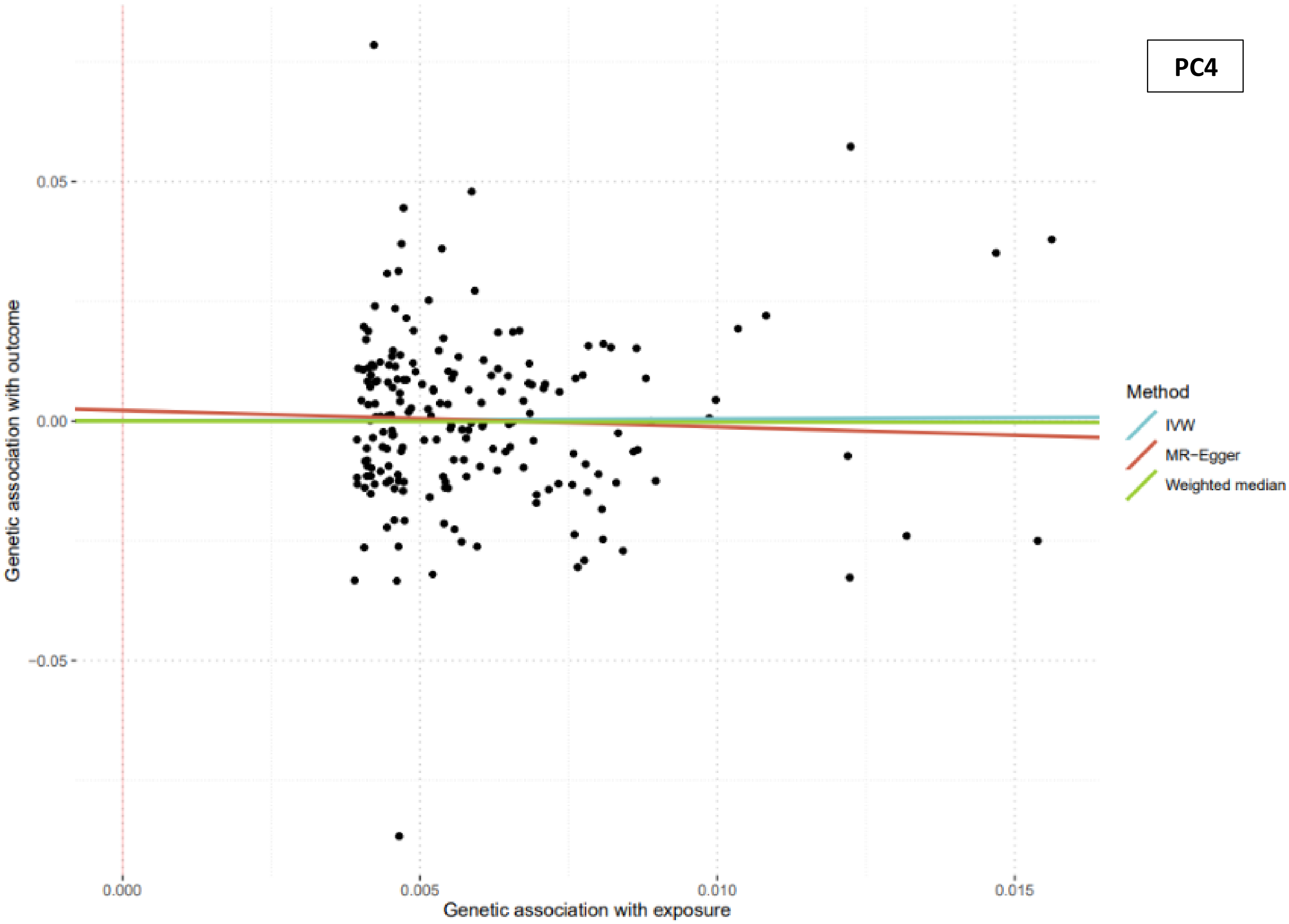

Figure S12. Scatter plot depicting the genetic associations of the body shape PC4 related instruments with the risk of overall colorectal cancer. PC, principal component.

	Adipose Tissue	Blood Vessel	Muscle - Skeletal	Uterus
Uterus				13
Muscle - Skeletal			68	12
Blood Vessel		79	52	13
Adipose Tissue	83	64	54	13

	Adipose Tissue	Blood Vessel	Breast - Mammary Tissue	Nerve - Tibial	Ovary	Uterus
Uterus						40
Ovary					65	29
Nerve - Tibial				267	61	39
Breast - Mammary Tissue			154	140	53	37
Blood Vessel		259	131	206	60	39
Adipose Tissue	278	203	148	217	59	38

	Blood Vessel	Nerve - Tibial	Ovary	Uterus
Uterus				89
Ovary			144	66
Nerve - Tibial		592	138	88
Blood Vessel	601	463	135	86

	Brain	Pituitary
Pituitary		110
Brain	204	90

Figure S13. Number of genetic variants included in the tissue-specific MR analysis that are shared between tissues of each body shape phenotype (PC1 – 4). PC1 in blue, PC2 in green, PC3 in red, and PC4 in yellow.

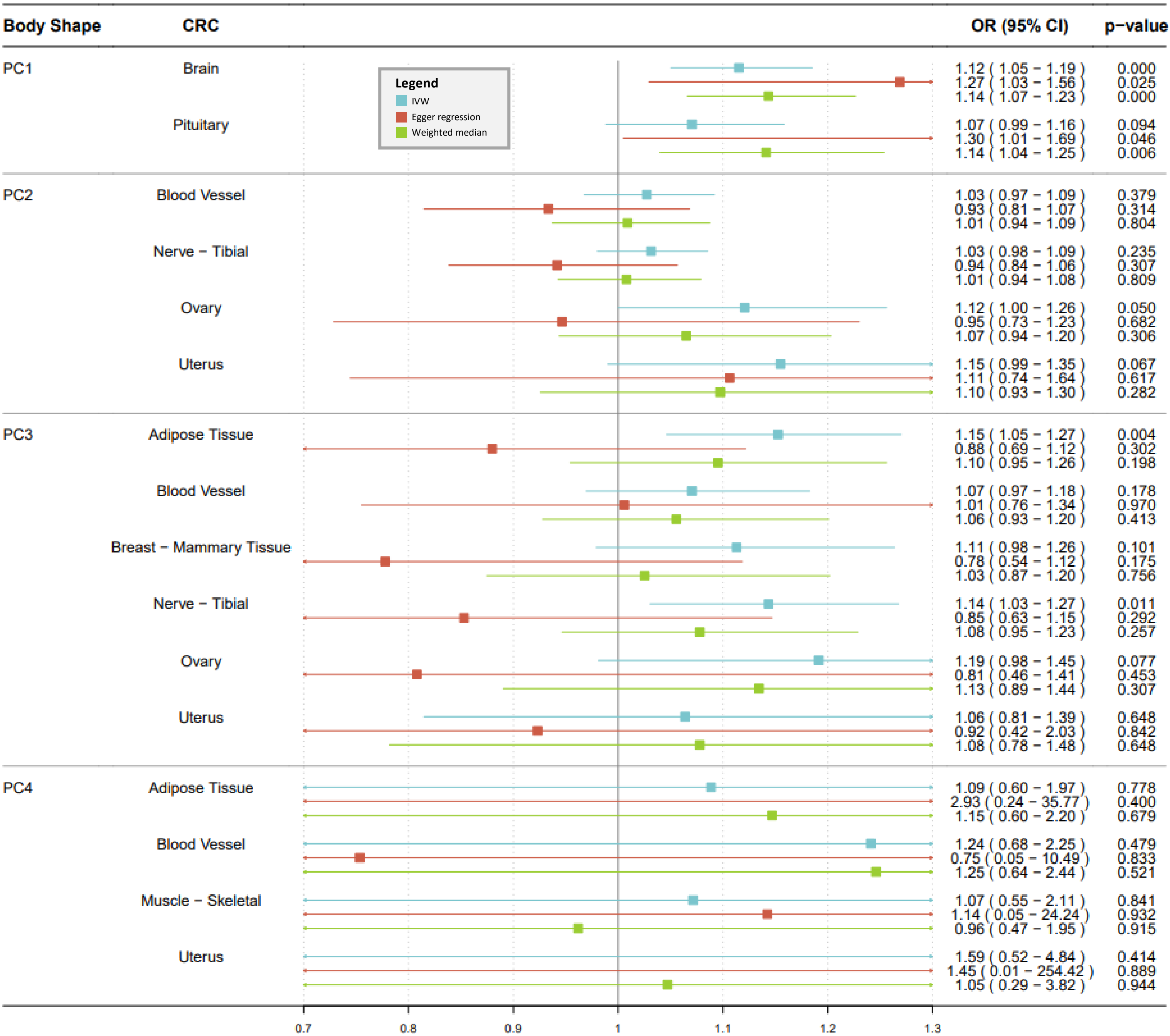

Figure S14. Mendelian randomization (MR) estimates between body shape phenotypes (PC1 - 4) grouped gene sets and overall colorectal cancer (CRC) risk using random-effects inverse-variance weighted (IVW) (blue), MR Egger regression (red) and weighted median (green) MR approaches. CI, confidence interval; OR, odds ratio; PC, principal component.

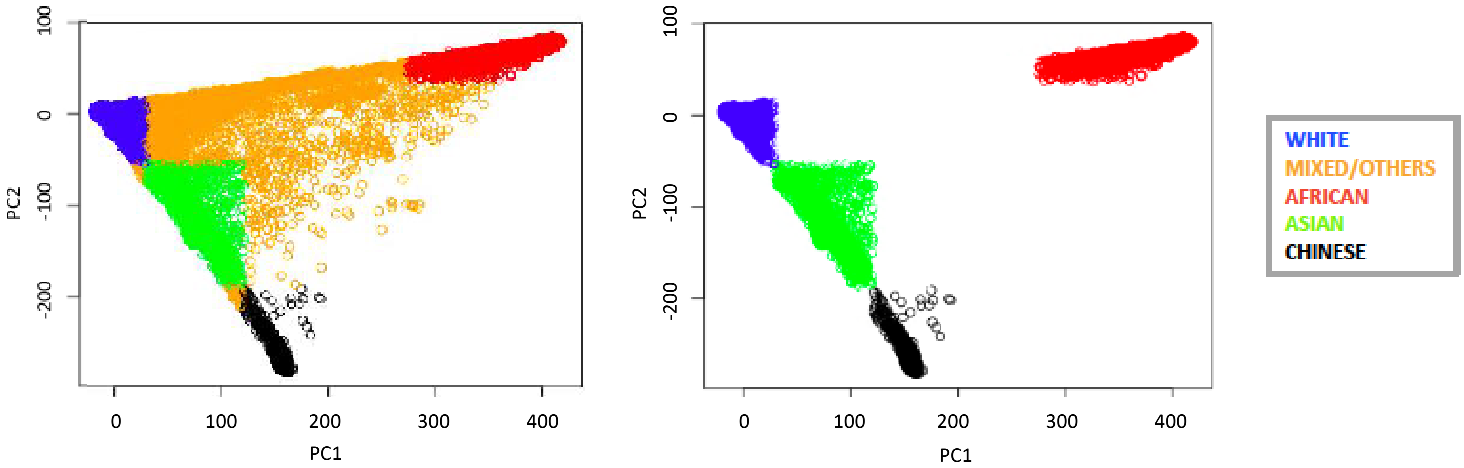

Figure S15. Clustering of individuals by ethnicity. Plot of genetic principal component (PC) 1 and PC2 for all the individuals included in the genetic analysis (n= 460,198). Based on self-reported information (White, other, mixed, or missing) and the clustering, 5 (White, mixed/others, African, Asian, and Chinese) and 4 (White, African, Asian, and Chinese) ethnic groups are represented in the left and right figures, respectively.

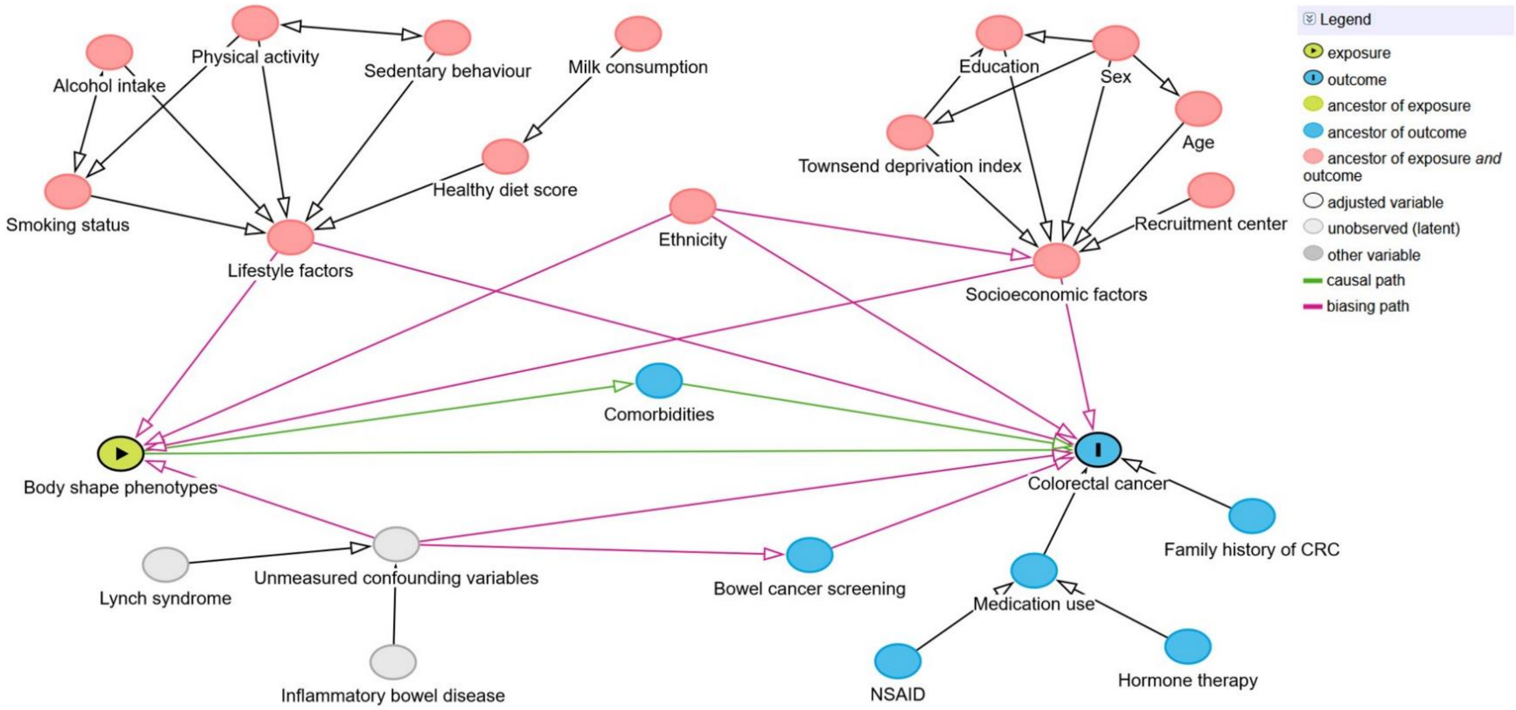

Figure S16. Extended directed acyclic graph (DAG) depicting the assumed causal relationship between body shape phenotypes and colorectal cancer risk with its confounding and mediating paths.

Socioeconomic factors: age, sex, recruitment center, the Townsend deprivation index, and education; ethnicity: White, Mixed, Asian/British Asian, Black/Black British, Chinese, other; lifestyle factors: tobacco smoking, physical activity, sedentary behavior, adherence to a healthy diet score, milk intake, alcohol intake frequency; medication use: nonsteroidal anti-inflammatory drugs (NSAID) and hormone therapy in postmenopausal women; bowel cancer screening, and family history of colorectal cancer (father and/or mother).

We did not adjust for comorbidities such as type 2 diabetes, because we assumed that in the pathway from body shapes to colorectal cancer this comorbidity would rather be a mediator than a confounder.

Unmeasured (known) confounders: inflammatory bowel syndrome and Lynch syndrome. Both phenotypes are difficult to diagnose clinically and data availability in the UK Biobank is 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 family history of colorectal cancer.

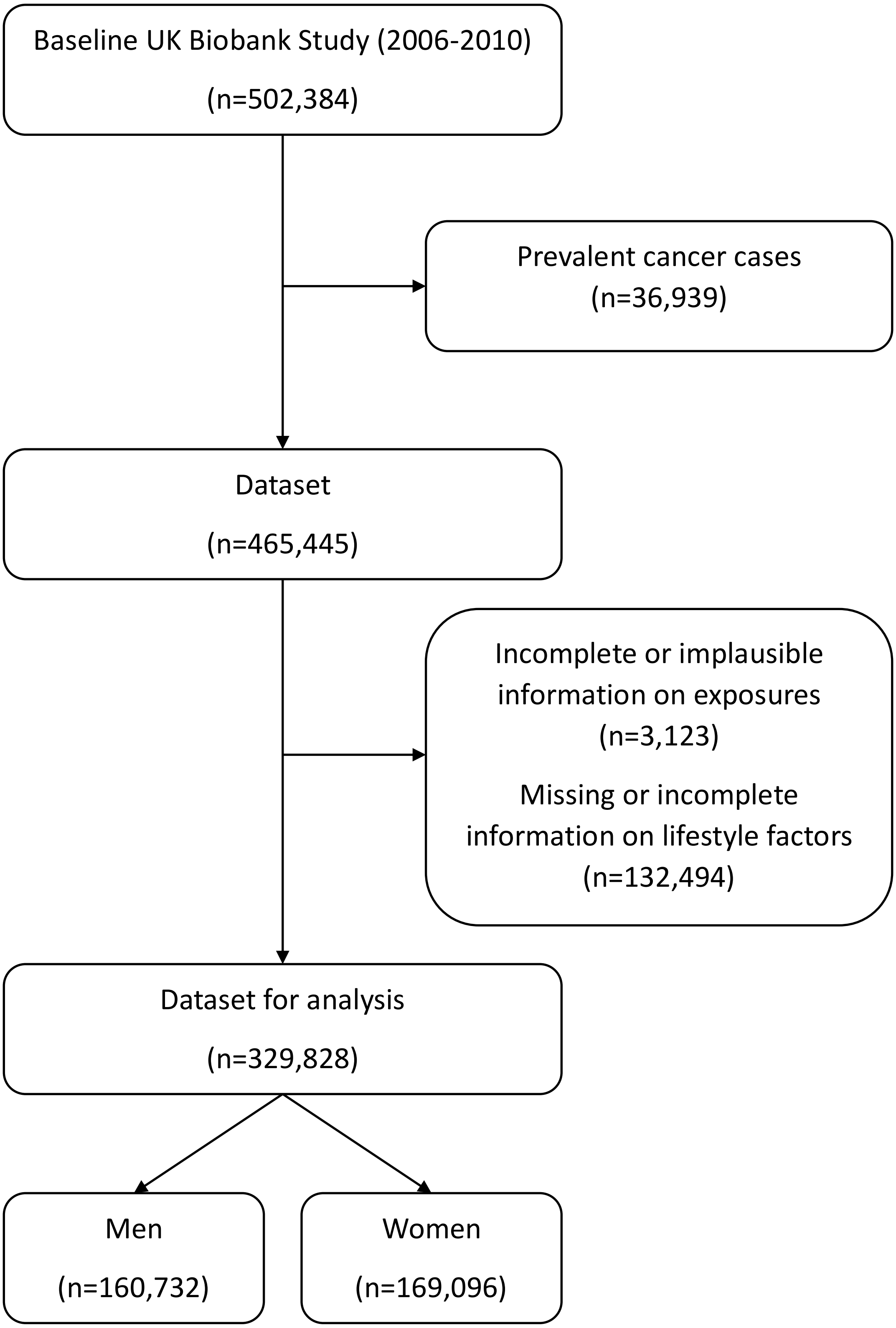

Figure S17. Flowchart of the UK Biobank study population.

Observational Analysis	PC1	PC2	PC3	PC4
Height	0.081	0.853	0.417	0.023
Weight	0.478	0.218	-0.078	0.509
BMI	0.471	-0.154	-0.288	0.471
HC	0.448	0.180	-0.404	-0.606
WC	0.488	-0.109	0.167	-0.388
WHR	0.323	-0.395	0.739	-0.020
Proportion of variance	0.662	0.194	0.123	0.020
Cumulative Proportion	0.662	0.856	0.979	0.999
Genetic Analysis	**PC1**	**PC2**	**PC3**	**PC4**
Height	0.072	0.865	0.399	0.020
Weight	0.478	0.217	-0.082	0.512
BMI	0.472	-0.152	-0.279	0.469
HC	0.448	0.172	-0.406	-0.604
WC	0.488	-0.101	0.168	-0.391
WHR	0.323	-0.376	0.751	-0.019
Proportion of variance	0.663	0.192	0.125	0.019
Cumulative Proportion	0.663	0.855	0.980	0.999

Table S1. Principal components (PC) loadings, proportion of variance and cumulative proportion for the individuals included in observational and genetic analyses. Sample size in observational analysis = 329,828 and in genetic analysis = 460,198 participants. BMI, body mass index; HC, hip circumference; WC, waist circumference; WHR, waist-to-hip ratio.

	PC1		PC2		PC3		PC4	
	Men	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Height	0.042	0.113	0.851	0.842	0.448	0.414	0.022	0.050
Weight	0.481	0.475	0.226	0.208	-0.086	-0.070	0.489	0.492
BMI	0.479	0.463	-0.106	-0.195	-0.272	-0.289	0.479	0.502
HC	0.457	0.440	0.181	0.197	-0.354	-0.444	-0.609	-0.584
WC	0.489	0.485	-0.110	-0.109	0.184	0.137	-0.399	-0.398
WHR	0.300	0.344	-0.410	0.398	0.747	0.724	0.028	-0.063

Table S2. Principal components (PC) loadings by sex (genetic analyses sample size). BMI, body mass index; HC, hip circumference; WC, waist circumference; WHR, waist-to-hip ratio.

	Men (n=160,732)					Women (n=169,096)				
	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
	Mean (SD) or %					Mean (SD) or %				
No. (%)	32,121 (20.0)	31,902 (19.8)	32,419 (20.2)	32,017 (19.9)	32,273 (20.1)	33,791 (20.0)	34,019 (20.1)	33,897 (20.0)	33,682 (19.9)	33,707 (19.9)
Age at recruitment (years)	56.5 (8.3)	56.5 (8.2)	56.4 (8.2)	56.4 (8.1)	56.0 (8.0)	55.1 (8.1)	55.3 (8.1)	56.0 (8.0)	56.2 (7.9)	55.6 (7.8)
Median follow-up time (years)	10.9	10.9	10.9	10.9	10.9	10.9	11.0	10.9	11.0	10.9
Anthropometric variables										
Weight (kg)	69.7 (5.7)	78.2 (4.3)	84.2 (4.4)	91.1 (4.9)	106.0 (11.5)	56.1 (4.6)	63.2 (3.9)	68.6 (4.2)	75.4 (4.9)	91.5 (12.0)
Height (cm)	173.6 (6.6)	175.2 (6.5)	176.1 (6.5)	177.0 (6.6)	178.0 (6.8)	161.5 (6.0)	162.8 (6.2)	163.1 (6.3)	163.2 (6.3)	163.5 (6.3)
Body mass index (kg/m^2)	23.2 (1.8)	25.5 (1.5)	27.2 (1.5)	29.1 (1.7)	33.5 (3.6)	21.5 (1.7)	23.9 (1.5)	25.8 (1.6)	28.3 (1.9)	34.3 (4.5)
Waist circumference (cm)	83.0 (4.7)	90.6 (3.1)	95.6 (3.1)	101.1 (3.4)	112.4 (8.2)	69.7 (3.8)	76.6 (3.1)	82.2 (3.4)	88.9 (4.1)	102.1 (8.8)
Hip circumference (cm)	95.5 (3.9)	99.8 (3.3)	102.6 (3.3)	105.8 (3.5)	113.0 (7.0)	92.5 (4.3)	97.7 (3.9)	101.3 (4.1)	105.9 (4.7)	117.1 (9.5)
Waist-to-hip ratio	0.9 (0.1)	0.9 (0.1)	0.9 (0.1)	1.0 (0.1)	1.0 (0.1)	0.8 (0.1)	0.8 (0.1)	0.8 (0.1)	0.8 (0.1)	0.9 (0.1)
Caucasian ethnicity (%)	94.2	95.0	95.7	96.2	96.5	95.7	96.1	95.7	95.0	93.9
Townsend deprivation index	-1.4 (3.1)	-1.6 (3.0)	-1.6 (2.90)	-1.6 (3.0)	-1.2 (3.1)	-1.5 (2.9)	-1.7 (2.9)	-1.7 (2.8)	-1.4 (2.9)	-0.9 (3.2)
Regular NSAID use (%)	21.4	25.1	27.2	30.7	36.3	20.1	22.4	24.8	27.5	32.1
Hormone replacement therapy (%)	0	0	0	0	0	32.8	35.2	37.7	39.0	36.0
Family history of CRC (%)	6.5	6.8	7.1	6.9	6.6	5.8	5.6	6.0	5.7	5.4
Bowel cancer screening (%)	31.6	32.2	32.5	32.5	31.1	28.1	28.1	29.5	29.9	28.9
Highest qualification (%)										
College/University degree	44.9	41.4	38.3	34.3	30.3	42.5	39.0	35.6	32.5	29.7
A-levels/AS-levels or equivalent	22.7	23.9	24.6	25.6	26.1	21.9	21.8	22.6	23.4	23.7
O-levels/GCSEs or equivalent	20.7	22.6	23.9	25.6	27.2	25.9	28.2	28.9	29.7	30.9
None (of the above)	11.8	12.1	13.2	14.5	16.4	9.8	11.0	12.9	14.5	15.6
MET (min/week)	3153.9 (2992.8)	2925.6 (2850.6)	2784.3 (2801.7)	2671.5 (2795.6)	2322.8 (2656.5)	2891.9 (2590.9)	2729.9 (2483.4)	2568.8 (2409.5)	2422.0 (2366.6)	2085.4 (2252.0)
Sedentary behavior (h/day)	4.5 (2.4)	4.9 (2.4)	5.2 (2.5)	5.5 (2.6)	6.1 (2.9)	3.8 (1.9)	4.0 (1.9)	4.3 (2.0)	4.6 (2.1)	5.0 (2.4)
Highest diet score (%)										
Lowest	0.3	0.2	0.3	0.3	0.4	0.0	0.0	0.1	0.1	0.1
Low	1.9	1.8	1.9	2.3	3.0	0.5	0.5	0.5	0.7	1.1
Low/medium	8.5	8.8	10.1	11.2	13.0	3.6	3.8	4.2	4.5	6.3
Medium	26.5	29.4	30.5	32.3	34.6	18.3	20.2	21.3	23.0	25.5
Medium/high	38.4	39.4	39.2	37.8	35.3	46.0	47.8	47.2	47.0	44.8
High	20.8	18.0	16.1	14.7	12.4	27.0	24.2	23.5	21.8	19.8
Highest	3.6	2.4	2.0	1.5	1.3	4.6	3.5	3.1	2.9	2.5
Milk intake >300ml/day (%)	74.1	75.0	73.3	72.9	70.8	68.8	71.8	73.4	72.5	70.7
Smoking status (%)										
Never	57.8	53.6	51.0	46.8	43.6	64.7	61.9	60.1	58.5	57.7
Previous	29.3	35.4	38.1	42.2	45.1	27.1	30.4	32.1	33.3	34.0
Current	12.9	11.0	10.9	11.0	11.3	8.3	7.8	7.8	8.2	8.4
Alcohol intake (%)										
Daily or almost daily	27.3	28.0	27.6	26.6	23.1	20.5	19.2	18.5	16.4	11.4
3-4 times per week	26.7	28.4	28.6	28.3	24.5	24.2	24.9	23.1	21.5	16.4
1-2 times per week	24.2	25.1	25.3	25.8	27.6	25.4	26.8	27.0	26.5	25.0
1-3 times per month	8.4	7.9	8.0	8.3	10.7	11.1	11.6	12.5	13.5	16.1
Special occasions only	7.0	5.8	5.6	6.0	8.0	11.0	10.9	12.0	14.0	20.0
Never	6.4	4.8	4.8	4.9	6.2	7.7	6.8	6.9	8.2	11.2

Table S3. Baseline characteristics of participants according to sex-specific quintiles of loadings of principal component 1 in the UK Biobank (observational analysis sample size). Sex-specific quintiles of loadings of principal component 1 were defined by their distribution (20th quantile, 40th quantile, 60th quantile, 80th quantile); for men: 20th quantile=-0.76, 40th quantile=-0.32, 60th quantile=0.10, 80th quantile=0.65; for women: 20th quantile=-0.91, 40th quantile=-0.45, 60th quantile=0.03, 80th quantile=0.71. CRC: colorectal cancer; NSAID: non-steroidal anti-inflammatory drug.

		PC1	PC2	PC3	PC4
Complete case					
	Cases (n)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
TOTAL (n=329,828)					
Colon	2443	**1.16 (1.11-1.21)**	0.97 (0.93-1.01)	**1.10 (1.06-1.14)**	0.99 (0.95-1.03)
Colon distal	966	**1.15 (1.07-1.23)**	0.97 (0.91-1.04)	**1.10 (1.03-1.17)**	0.96 (0.90-1.03)
Colon proximal	1370	**1.17 (1.11-1.24)**	0.97 (0.91-1.02)	**1.11 (1.05-1.17)**	1.00 (0.94-1.05)
Colorectal	3728	**1.13 (1.09-1.17)**	0.98 (0.94-1.01)	**1.09 (1.05-1.13)**	0.99 (0.96-1.03)
Rectum	1285	**1.07 (1.01-1.14)**	0.99 (0.93-1.04)	**1.07 (1.01-1.14)**	1.00 (0.95-1.06)
MEN (n=160,732)					
Colorectal	2239	**1.17 (1.11-1.23)**	0.97 (0.93-1.02)	**1.09 (1.04-1.15)**	0.98 (0.94-1.02)
WOMEN (n=169,096)					
Colorectal	1489	**1.09 (1.03-1.15)**	0.97 (0.92-1.03)	**1.08 (1.03-1.13)**	1.00 (0.94-1.05)
Never smokers					
TOTAL (n=183,679)					
Colon	1207	**1.16 (1.09-1.23)**	0.96 (0.91-1.02)	**1.11 (1.05-1.17)**	1.01 (0.96-1.07)
Colon distal	468	**1.12 (1.01-1.23)**	0.99 (0.90-1.09)	1.05 (0.96-1.15)	1.01 (0.92-1.11)
Colon proximal	683	**1.19 (1.10-1.29)**	0.94 (0.87-1.02)	**1.14 (1.06-1.23)**	1.00 (0.93-1.08)
Colorectal	1787	**1.13 (1.07-1.19)**	0.95 (0.90-1.00)	**1.09 (1.04-1.14)**	1.00 (0.95-1.05)
Rectum	580	1.08 (0.98-1.18)	0.93 (0.85-1.01)	1.06 (0.97-1.15)	0.96 (0.88-1.05)
MEN (n=81,263)					
Colorectal	949	**1.20 (1.11-1.29)**	0.91 (0.86-0.98)	**1.12 (1.04-1.20)**	0.97 (0.91-1.04)
WOMEN (n=102,416)					
Colorectal	838	1.07 (1.00-1.15)	0.98 (0.92-1.06)	1.07 (1.00-1.14)	1.00 (0.93-1.08)
Follow-up time > 2 years					
TOTAL (n=324,431)					
Colon	2087	**1.17 (1.12-1.23)**	0.97 (0.93-1.02)	**1.10 (1.05-1.15)**	1.00 (0.95-1.04)
Colon distal	802	**1.14 (1.06-1.23)**	0.96 (0.89-1.03)	**1.10 (1.02-1.18)**	0.97 (0.90-1.04)
Colon proximal	1206	**1.18 (1.11-1.26)**	0.97 (0.91-1.03)	**1.10 (1.04-1.17)**	1.02 (0.96-1.08)
Colorectal	3173	**1.14 (1.10-1.18)**	0.97 (0.94-1.01)	**1.09 (1.05-1.13)**	1.00 (0.97-1.04)
Rectum	1086	**1.08 (1.01-1.15)**	0.99 (0.93-1.05)	**1.07 (1.00-1.14)**	1.01 (0.95-1.08)
MEN (n=157,710)					
Colorectal	1875	**1.17 (1.11-1.23)**	0.98 (0.93-1.02)	**1.09 (1.03-1.15)**	0.98 (0.94-1.03)
WOMEN (n=166,721)					
Colorectal	1298	**1.11 (1.05-1.18)**	0.97 (0.91-1.03)	**1.08 (1.03-1.14)**	1.01 (0.96-1.07)
Crude model					
TOTAL (n=329,828)					
Colon	2443	**1.18 (1.13-1.23)**	0.98 (0.94-1.02)	**1.11 (1.06-1.15)**	0.99 (0.95-1.03)
Colon distal	966	**1.16 (1.09-1.24)**	0.98 (0.92-1.04)	**1.11 (1.04-1.18)**	0.96 (0.90-1.03)
Colon proximal	1370	**1.19 (1.12-1.25)**	0.98 (0.93-1.03)	**1.12 (1.06-1.18)**	0.99 (0.94-1.05)
Colorectal	3728	**1.14 (1.10-1.18)**	0.98 (0.95-1.02)	**1.10 (1.06-1.14)**	0.99 (0.96-1.02)
Rectum	1285	**1.07 (1.01-1.14)**	0.99 (0.94-1.05)	**1.09 (1.02-1.15)**	1.00 (0.95-1.06)
MEN (n=160,732)					
Colorectal	2239	**1.18 (1.13-1.24)**	0.98 (0.94-1.02)	**1.11 (1.05-1.16)**	0.97 (0.93-1.02)
WOMEN (n=169,096)					
Colorectal	1489	**1.10 (1.04-1.16)**	0.99 (0.93-1.04)	**1.08 (1.03-1.14)**	1.00 (0.95-1.05)

Table S4. Associations between body shape phenotypes (PC1 - 4) per 1 SD increment and colorectal cancer (CRC) incidence in the complete case dataset (n=329,828), in never smokers (n=183,679), after further excluding the initial 2 years of follow-up (n=324,431), and as a crude model (n=329,828). Hazard ratios per 1 SD increment in each PC from Cox proportional hazards regression using age as the underlying time metric. Stratified by sex, age in 5-year categories, and center. Adjusted for ethnicity, Townsend deprivation index, intake of non-steroidal anti-inflammatory drugs, education, smoking status, alcohol status and intake frequency, MET Min/week, sedentary behavior, healthy diet score, milk intake, hormone replacement therapy, family history of CRC, bowel cancer screening. All four principal components were mutually adjusted. Crude hazard ratios per 1 SD increment in each PC from Cox proportional hazards regression using age as the underlying time metric in the complete case dataset (n=329,828). Stratified by sex, age in 5-year categories, and center. All four principal components were mutually adjusted. Bold font indicates statistical significance (p-value < 0.05). PC, principal component.

Outcome	Exposure	Heterogeneity			Egger intercept		
		Cochran's Q	df	p-value	Estimate	95% CI	p-value
Overall	**PC1**	999.099	671	< 0.001	0.000	-0.003 , 0.003	0.898
	PC2	3043.754	1773	< 0.001	0.001	-0.001 , 0.002	0.382
	PC3	1139.40	733	< 0.001	0.000	-0.003 , 0.003	0.892
	PC4	450.8924	192	< 0.001	0.002	-0.007 , 0.012	0.649
Overall (Men)	**PC1**	894.842	671	< 0.001	0.000	-0.004 , 0.004	0.923
	PC2	2408.643	1773	< 0.001	0.001	-0.001 , 0.003	0.354
	PC3	960.500	733	< 0.001	0.000	-0.003 , 0.004	0.832
	PC4	326.4463	192	< 0.001	0.004	-0.007 , 0.016	0.487
Overall (Women)	**PC1**	778.338	671	0.003	-0.001	-0.004 , 0.003	0.764
	PC2	2587.901	1773	< 0.001	0.001	-0.001 , 0.003	0.554
	PC3	888.611	733	< 0.001	-0.001	-0.004 , 0.003	0.763
	PC4	325.8687	192	< 0.001	0.001	-0.011 , 0.012	0.903
Colon	**PC1**	950.152	671	< 0.001	0.000	-0.004 , 0.003	0.954
	PC2	2789.540	1773	< 0.001	0.000	-0.002 , 0.002	0.738
	PC3	1025.401	733	< 0.001	-0.001	-0.004 , 0.002	0.541
	PC4	373.7163	192	< 0.001	0.002	-0.008 , 0.013	0.668
Distal colon	**PC1**	866.015	671	< 0.001	0.000	-0.004 , 0.004	0.964
	PC2	2424.078	1773	< 0.001	0.001	-0.002 , 0.003	0.513
	PC3	919.797	733	< 0.001	-0.001	-0.005 , 0.003	0.550
	PC4	317.9822	192	< 0.001	-0.002	-0.014 , 0.011	0.766
Proximal colon	**PC1**	898.1401	671	< 0.001	0.000	-0.004 , 0.004	0.948
	PC2	2493.266	1773	< 0.001	0.000	-0.003 , 0.002	0.773
	PC3	936.479	733	< 0.001	0.000	-0.004 , 0.003	0.878
	PC4	289.055	192	< 0.001	0.007	-0.004 , 0.019	0.221
Rectal	**PC1**	803.636	671	< 0.001	0.000	-0.004 , 0.005	0.871
	PC2	2372.954	1773	< 0.001	0.001	-0.001 , 0.004	0.248
	PC3	921.793	733	< 0.001	-0.002	-0.006 , 0.002	0.392
	PC4	376.2022	192	< 0.001	0.000	-0.014 , 0.014	0.995

Table S5. Heterogeneity test and Egger intercept results for body shape phenotypes (PC1 - 4) and overall colorectal cancer risk, subsites and by sex. CI, confidence interval; PC, principal component.

PC	Tissue	Heterogeneity			Egger intercept		
		Cochran's Q	df	p-value	Estimate	95% CI	p-value
PC1	**Brain**	394.3695	203	< 0.001	-0.005	-0.012 , 0.003	0.207
	Pituitary	205.2431	109	< 0.001	-0.007	-0.017 , 0.002	0.118
PC2	**Blood vessel**	1169.7420	600	< 0.001	0.002	0.000 , 0.005	0.089
	Nerve - Tibial	1092.7574	591	< 0.001	0.003	-0.000 , 0.005	0.080
	Ovary	341.2845	143	< 0.001	0.005	-0.002 , 0.011	0.161
	Uterus	212.7641	88	< 0.001	0.001	-0.008 , 0.010	0.817
PC3	**Adipose tissue**	453.8644	277	< 0.001	0.005	0.001 , 0.010	0.018
	Blood Vessel	393.5795	258	< 0.001	0.001	-0.004 , 0.006	0.648
	Breast	241.6158	153	< 0.001	0.007	0.000 , 0.013	0.039
	Nerve - Tibial	440.7192	266	< 0.001	0.005	0.000 , 0.011	0.039
	Ovary	100.4234	64	0.0025	0.007	-0.002 , 0.017	0.145
	Uterus	67.2299	39	0.0033	0.003	-0.011 , 0.016	0.707
PC4	**Adipose tissue**	167.2578	82	< 0.001	-0.006	-0.020 , 0.008	0.425
	Blood Vessel	147.6044	78	< 0.001	0.003	-0.012 , 0.017	0.703
	Muscle	146.9878	67	< 0.001	0.000	-0.017 , 0.017	0.966
	Uterus	16.8309	12	0.1561	0.001	-0.030 , 0.031	0.971

Table S6. Heterogeneity test and Egger intercept results for body shape phenotypes (PC1 - 4) grouped gene sets and overall colorectal cancer risk. CI, confidence interval; PC, principal component.

	PC1				PC2			
	5th	95th	Difference	SD-unit difference	5th	95th	Difference	SD-unit difference
BMI (kg/m^2)	20.6	39.5	18.9	3.9	29.2	26.1	-3.1	-0.6
Weight (kg)	56.1	112.3	56.2	3.5	72.6	86.6	14.0	0.9
Height (cm)	165.0	168.7	3.7	0.4	157.4	181.8	24.4	2.7
WHR	0.8	0.9	0.1	1.0	0.9	0.8	-0.1	-1.0
WC (cm)	70.2	117.8	47.6	3.5	94.4	88.9	-5.5	-0.4
HC (cm)	90.2	125.8	35.6	3.9	100.1	107.5	7.4	0.8
	PC3				PC4			
	5th	95th	Difference	SD-unit difference	5th	95th	Difference	SD-unit difference
BMI (kg/m^2)	33.1	26.9	-6.2	-1.3	28.2	31.2	3.0	0.6
Weight (kg)	86.8	79.3	-7.5	-0.5	79.2	90.0	10.8	0.7
Height (cm)	161.8	171.3	9.5	1.0	168.1	169.4	1.3	0.1
WHR	0.8	1.0	0.2	2.0	0.9	0.9	0.0	0.0
WC (cm)	92.6	97.2	4.6	0.3	99.1	92.4	-6.7	-0.5
HC (cm)	116.4	99.9	-16.5	-1.8	111.3	102.5	-8.8	-1.0

Table S7. Mean values and the difference in standard deviation (SD) units of each anthropometric trait for study participants in the top and bottom 5% of each body shape phenotypes (PC1 - 4). 1 SD unit for BMI in the study population corresponded to 4.8 kg/m^2, weight = 15.9 kg, height = 9.2 cm, WHR = 0.1, WC = 13.5 cm, HC = 9.2 cm. For instance, the difference between 95th and 5th percentile in PC1 corresponds to 3.9 SD BMI-units difference (18.9 / 4.8 = 3.9). BMI, body mass index; HC, hip circumference; PC, principal component; WC, waist circumference; WHR, waist-to-hip ratio.

	Men	Women
PC1	0.015 (1.86)	-0.028 (2.08)
PC2	0.042 (1.09)	0.043 (1.03)
PC3	0.008 (0.76)	-0.001 (0.93)
PC4	0.003 (0.34)	0.002 (0.34)

Table S8. Mean and standard deviations of principal component (PC) scores for each body shape phenotypes (PC1 - 4) and by sex. PC, principal component.

General tissue	Specific tissue	Sample size
Adipose Tissue	Subcutaneous	581
	Visceral (Omentum)	469
Blood Vessel	Aorta	387
	Coronary	213
	Tibial	584
Brain	Amygdala	129
	Anterior cingulate cortex (BA24)	147
	Caudate (basal ganglia)	194
	Cerebellar Hemisphere	175
	Cerebellum	209
	Cortex	205
	Front Cortex (BA9)	175
	Hippocampus	165
	Hypothalamus	170
	Nucleus accumbens (basal ganglia)	202
	Putamen (basal ganglia)	170
	Spinal cord (cervical c-1)	126
	Substantia nigra	114
Breast - Mammary Tissue		396
Muscle - Skeletal		706
Nerve -Tibial		532
Ovary		167
Pituitary		237
Uterus		129

Table S9. GTEx data sample sizes by specific tissue.

Body shape	General tissue	Specific tissue	Num. genetic variants
PC1	Brain	Amygdala	40
		Anterior cingulate cortex (BA24)	55
		Caudate (basal ganglia)	87
		Cerebellar Hemisphere	113
		Cerebellum	135
		Cortex	105
		Front Cortex (BA9)	89
		Hippocampus	58
		Hypothalamus	56
		Nucleus accumbens (basal ganglia)	84
		Putamen (basal ganglia)	74
		Spinal cord (cervical c-1)	47
		Substantia nigra	30
	Pituitary		110
PC2	Blood Vessel	Aorta	418
		Coronary	181
		Tibial	550
	Nerve-Tibial		592
	Ovary		144
	Uterus		89
PC3	Adipose Tissue	Subcutaneous	258
		Visceral (Omentum)	180
	Blood Vessel	Aorta	181
		Coronary	75
		Tibial	231
	Breast – Mammary Tissue		154
	Nerve-Tibial		267
	Ovary		65
	Uterus		40
PC4	Adipose Tissue	Subcutaneous	79
		Visceral (Omentum)	55
	Blood Vessel	Aorta	54
		Coronary	28
		Tibial	73
	Muscle – Skeletal		68
	Uterus		13

Table S10. Genetic variants mapped in each specific tissue included as genetic instruments in the specific-tissue MR analysis. PC, principal component.

Supplementary data legend

Data S1. Summary statistics for the genetic association with body shapes PC1-4.

Supplementary Text. GECCO Funding and Acknowledgments

Funding

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, R01 201407). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN268201700006I and HHSN268201200008I . This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685.

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

CLUE II funding was from the National Cancer Institute (U01 CA086308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG018033), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

Maryland Cancer Registry (MCR)

Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

ColoCare: This work was supported by the National Institutes of Health (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueiredo/Colditz, 2P30CA015704- 40 (Gilliland), R01 CA207371 (Ulrich/Li)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143237 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN268201200008I. Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of Health Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to SG), and through generous support from the Ontario Ministry of Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part through NCI/NIH awards U01/U24 CA074794 and R01 CA076366 (to PAN). The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alpe d'Huzes and the Dutch Cancer Society (UM 2012–5653, UW 2013-5927, UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The Nqplus study is sponsored by a ZonMW investment grant (98-10030); by PREVIEW, the project PREVention of diabetes through lifestyle intervention and population studies in Europe and around the World (PREVIEW) project which received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant no. 312057; by funds from TI Food and Nutrition (cardiovascular health theme), a public–private partnership on precompetitive research in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI Healthy Diet for a Healthy Life.

COLO2&3: National Institutes of Health (R01 CA060987).

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA081488, P30 CA014089, R01 CA197350; P01 CA196569; R01 CA201407; R01 CA242218), National Institutes of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678) and a generous gift from Daniel and Maryann Fong.

CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829675, to Andrea Gsur), the "Herzfelder'sche Familienstiftung" (grant to Andrea Gsur) and was supported by COST Action BM1206.

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. The study protocol was approved by the institutional review boards of Emory University, and those of participating registries as required.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. We thank CERCA Programme, Generalitat de Catalunya for institutional support.

Czech Republic CCS: This work was supported by the Grant Agency of the Czech Republic (21-04607X, 20-03997S), by the Grant Agency of the Ministry of Health of the Czech Republic (grants AZV NU21-07-00247 and AZV NU21-03-00506), and Charles University Research Fund (Cooperation 43-Surgical disciplines)

DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: National Institutes of Health (R01 CA048998 to M. L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN Grant (U01-CA152753).

EPIC: The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam- Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and and Region Skåne and Region Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom).

EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02368, P111/00219, PI11/00681, PI14/00173, PI14/00230, PI17/00509, 17/00878, PI20/00113, PI20/00226, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economia y Competitividad (SAF07-64873, SAF 2010-19273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya) and COST Action BM1206 and CA17118. CIBERehd is funded by the Instituto de Salud Carlos III.

ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid.

Harvard cohorts: HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993, and R35 CA197735), and PHS by the National Institutes of Health (R01 CA042182).

Hawaii Adenoma Study: NCI grants R01 CA072520.

HCES-CRC: the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital, HCRI15011-1).

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. BMLynch was supported by MCRF18005 from the Victorian Cancer Agency.

MEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a generous gift from Daniel and Maryann Fong.

MSKCC: The work at Sloan Kettering in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. Moffitt: This work was supported by funding from the National Institutes of Health (grant numbers R01 CA189184, P30 CA076292), Florida Department of Health Bankhead-Coley Grant 09BN-13, and the University of South Florida Oehler Foundation. Moffitt contributions were supported in part by the Total Cancer Care Initiative, Collaborative Data Services Core, and Tissue Core at the H. Lee Moffitt Cancer Center & Research Institute, a National Cancer Institute-designated Comprehensive Cancer Center (grant number P30 CA076292).

NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA066635 and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, U.S. Department of Health and Human Serivces (U01 CA074783); and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

NSHDS: The research was supported by Biobank Sweden through funding from the Swedish Research Council (VR 2017-00650, VR 2017-01737), the Swedish Cancer Society (CAN 2017/581), Region Västerbotten (VLL-841671, VLL-833291), Knut and Alice Wallenberg Foundation (VLL-765961), and the Lion’s Cancer Research Foundation (several grants) and Insamlingsstiftelsen, both at Umeå University.

OSUMC: OCCPI funding was provided by Pelotonia and HNPCC funding was provided by the NCI (CA016058 and CA067941).

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.

SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA037429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SMS and REACHS: This work was supported by the National Cancer Institute (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the National Institutes of Health (grant KL2 TR000421 to A.N.B.-H.)

The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish research council; K2015-55X-22674-01-4, K2008-55X-20157-03-3, K2006-72X-20157-01-2 and the Stockholm County Council (ALF project).

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute´s Distinguished Professor Award to Alicja Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614

VITAL: National Institutes of Health (K05 CA154337).

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005

Acknowledgments

ASTERISK: We are very grateful to those, without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

CCFR: The Colon CFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the U.S. National Cancer Institute, without which this important registry would not exist. The authors would like to thank the study participants and staff of the Seattle Colon Cancer Family Registry and the Hormones and Colon Cancer study (CORE Studies).

CLUE II: We thank the participants of Clue II and appreciate the continued efforts of the staff at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study. Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

COLON and NQplus: the authors would like to thank the COLON and NQplus investigators at Wageningen University & Research and the involved clinicians in the participating hospitals.

CORSA: We kindly thank all individuals who agreed to participate in the CORSA study. Furthermore, we thank all cooperating physicians and students and the Biobank Graz of the Medical University of Graz.

CPS-II: The authors express sincere appreciation to all Cancer Prevention Study-II (and/or Cancer Prevention Study-3) participants, and to each member of the study and biospecimen management group. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries and cancer registries supported by the National Cancer Institute's Surveillance Epidemiology and End Results Program. The authors assume full responsibility for all analyses and interpretation of results. The views expressed here are those of the authors and do not necessarily represent the American Cancer Society or the American Cancer Society – Cancer Action Network.

Czech Republic CCS: We are thankful to all clinicians in major hospitals in the Czech Republic, without whom the study would not be practicable. We are also sincerely grateful to all patients participating in this study.

DACHS: We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance.

EDRN: We acknowledge all contributors to the development of the resource at University of Pittsburgh School of Medicine, Department of Gastroenterology, Department of Pathology, Hepatology and Nutrition and Biomedical Informatics.

EPIC: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

EPICOLON: We are sincerely grateful to all patients participating in this study who were recruited as part of the EPICOLON project. We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clínic–IDIBAPS and Biobanco Vasco for the availability of the samples. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Harvard cohorts: The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We acknowledge Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital as home of the NHS. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming. The authors assume full responsibility for analyses and interpretation of these data.

Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.

LCCS: We acknowledge the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith and Emma Northwood in conducting this study.

NCCCS I & II: We would like to thank the study participants, and the NC Colorectal Cancer Study staff.

NSHDS investigators thank the Västerbotten Intervention Programme, the Northern Sweden MONICA study, the Biobank Research Unit at Umeå University and Biobanken Norr at Region Västerbotten for providing data and samples and acknowledge the contribution from Biobank Sweden, supported by the Swedish Research Council.

PLCO: The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible.

Cancer incidence data have been provided by the District of Columbia Cancer Registry, Georgia Cancer Registry, Hawaii Cancer Registry, Minnesota Cancer Surveillance System, Missouri Cancer Registry, Nevada Central Cancer Registry, Pennsylvania Cancer Registry, Texas Cancer Registry, Virginia Cancer Registry, and Wisconsin Cancer Reporting System. All are supported in part by funds from the Center for Disease Control and Prevention, National Program for Central Registries, local states or by the National Cancer Institute, Surveillance, Epidemiology, and End Results program. The results reported here and the conclusions derived are the sole responsibility of the authors.

SEARCH: We thank the SEARCH team

SELECT: We thank the research and clinical staff at the sites that participated on SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf