Supplementary Materials

Supplementary methods: creating a polygenic obesity risk score

The majority of SNPs (24/34) were available on the Affymetrix 6.0 GeneChip: rs9939609 (*FTO* (12)), rs2867125 (*TMEM18* (12)), rs571312 (*MC4R* (12)), rs10938397 (*GNPDA2* (12)), rs10767664 (*BDNF* (12)), rs2815752 (*NEGR1* (12)), rs7359397 (*SH2B1* (12)) rs3817334 (*MTCH2* (12)), rs29941 (*KCTD15* (12)), rs543874 (*SEC16B* (12)), rs987237 (*TFAP2B* (12)), rs7138803 (*FAIM2* (12)), rs10150332 (*NRXN3* (12)), rs713586 (*RBJ; POMC* (12)), rs12444979 (*GPRC5B* (12)), rs2241423 (*MAP2K5* (12)), rs1514175 (*TNNI3K* (12)), rs10968576 (*LRRN6C* (12)), rs887912 (*FANCL* (12)), rs13078807 (*CADM2* (12)), rs1555543 (*PTBP2* (12)), rs206936 (*NUDT3* (12)), rs9568856 (*OLFM4* (11)), rs9299 (*HOXB5* (11)). It was possible to index another four using proxy SNPs that were in high linkage disequilibrium with the original SNP ($R^2 > 0.9$): rs2112347 (*FLJ35779*) was indexed using rs3797580 ($R^2 = 1$); rs4836133 (*ZNF608*) was indexed using rs6864049 ($R^2 = 1$); rs4929949 (*RPL27A*) was indexed using rs9300093 ($R^2 = 0.97$); rs3810291 (*TMEM160*) was indexed using rs7250850 ($R^2 = 1$). Six of the 34 SNPs could not be measured directly or reliably tagged using proxy SNPs: rs2890652 (LRP1B), rs9816226 (ETV5), rs13107325 (SLC39A8), rs4771122 (MTIF3), rs11847697 (PRKD1), and rs2287019 (QPCTL).

Supplementary Table 1. Bivariate twin model-fitting results (with standard errors) for BMI-SDS between ages 4 and 10 years

variables	A				C				E				N*
	V(G)_tr1	V(G)_tr2	C(G)_tr12	r_G	V(c)_tr1	V(c)_tr2	C(c)_tr12	r_C	V(e)_tr1	V(e)_tr2	C(e)_tr12	r_E	
BMISDS 4 to 10	.43(.04)	.82(.04)	.34(.04)	.58(.05)	.41(.04)	.06(.04)	.01(.04)	.08(.51)	.16(.01)	.12(.01)	.04(.01)	.31(.04)	5090

Annotation: V(G) – proportion of the variance explained by genetic factors for trait 1 and trait 2 (tr1, tr2); C(G) –the raw covariance between trait 1 and 2 explained by genetic factors; V(c) – proportion of the variance explained by shared environment for trait 1 and trait 2 (tr1, tr2); C(c) –the raw covariance between trait 1 and 2 explained by shared environment; V(e) – proportion of the variance explained by non-shared environment for trait 1 and trait 2 (tr1, tr2); C(e) –the raw covariance between trait 1 and 2 explained by non-shared environment; rG – genetic correlation; rC – correlation of shared environmental factors; rE – correlation of non-shared environmental factors; N – number of individual twin children with data for either trait 1 or trait 2; values in parentheses are standard errors.

*OpenMx twin model-fitting incorporates full-information maximum likelihood that uses the full sample where at least one sibling has available data. The sample included 2556 children with genotyping data and BMI-SDS at age 4 or 10 years, along with 2534 co-twins who had BMI-SDS at age 4 or 10 years.

variables	A						E				Vp_tr1	Vp_tr2	n_tr1*	n_tr2*
	V(G)_tr1	V(G)_tr2	C(G)_tr12	V(G)/Vp_tr1	V(G)/Vp_tr2	r_G	V(e)_tr1	V(e)_tr2	C(e)_tr12	r_E**				
BMISDS 4 to 10	.20(.21)	.29(.14)	.16(.13)	.20(.21)	.29(.14)	.66(.48)	.81(.21)	.71(.14)	.25(.13)	.34(.15)	1.00(.04)	1.00(.03)	1419	2268

Annotation: V(G) – variance explained by genetic factors for trait 1 and trait 2 (tr1, tr2); C(G) – raw covariance between trait 1 and 2 explained by genetic factors; V(e) – residual variance for trait 1 and trait 2; C(e) – raw residual covariance between trait 1 and trait 2; Vp – phenotypic variance for trait 1 and trait 2; V(G) / Vp – proportion of the phenotypic variance explained by genetic factors for trait 1 and trait 2; r_G – genetic correlation between trait 1 and trait 2; logL – log likelihood estimation of the model; n – number of individuals with data for trait 1 and trait 2; values in parentheses are standard errors.

*GCTA incorporates full-information maximum likelihood that uses the full sample of 2556 individuals with data on trait 1 or trait 2. However, the variance estimates for each trait are based on individuals with data for that trait, the last two columns are sample sizes with data present at each age.

** The current version of GCTA does not report the environmental correlation or its standard error. The environmental correlation was derived here from the GCTA estimates using the following algorithm: C(e)_tr12 / (√V(e)_tr1 * √V(e)_tr2), whereas the standard error was calculated using: Var(re) = re * re * (VarVe1/(4*Ve1*Ve1) + VarVe2/(4*Ve2*Ve2) + VarCe/(Ce*Ce) + CovVe1Ve2/(2*Ve1*Ve2) - CovVe1Ce/(Ve1*Ce) - CovVe2Ce/(Ve2*Ce)); SE(re) = sqrt[Var(re)], where re is the environmental correlation, Ve1 is the residual variance for trait 1, Ce is the residual covariance between two traits, VarVe1 is the sampling variance for Ve1 (residual variance for trait 1), VarCe is the sampling variance for Ce, CovVe1Ve2 is the sampling covariance between Ve1 and Ve2, and CovVe1Ce is the sampling covariance between Ve1 and Ce.

Supplementary Figure 1. Sampling distribution for R^2 at age 4 (a), at age 10 (b), and for the difference in R^2 between ages 4 and 10 (c)