Genetics and life course epidemiology of cardiometabolic disease: towards personalized medicine

Ibi, D.

Genome-wide association analysis of body weight trajectories over the life course implicates the *TOMM40-APOE* locus in late life weight loss

Dorina Ibi, Martijn Dolle, Monique Verschuren, Ko Willems van Dijk, Jan Bert van Klinken

In preparation for submission.

Multiple genetic variants have been associated with obesity. However, little is known about the effect of these variants on bodyweight (BW) development during the different stages of adult life. We performed genome-wide association analyses on longitudinal measures of BW and on BW at specific age groups from 30-70 years, as determined using a mixed-model approach in a Dutch population of 4619 men and women. After filtering for known adult BMI loci, we found two loci with age-dependent effects on BW development. The minor allele of rs2075650 (*TOMM40/APOE*) was associated with progressive weight loss starting at the age of 50 (rs2075650-G: BW-slope beta=-3.0, p=4.9E-07). Interestingly, this variant is also strongly associated with Alzheimer's disease. In addition, rs543874 (*SEC16B*) was associated with a sudden weight increase during midlife (rs543874-G: BW 50 yrs beta=1.3, p=8.8E-06), with more pronounced effects in women. These data indicate that some genetic loci may have a variable effect on BW over the life course of an individual.

The obesity pandemic has put an increasing burden on healthcare systems due to comorbidities such as cardiovascular disease, type 2 diabetes, cancer and psychological disturbances [1,2] . To unravel the genetic architecture of obesity, anthropomorphic measurements such as body max index (BMI), fat mass, waist-to-hip ratio and bodyweight (BW) have been investigated in ever larger GWAS consortia, resulting in over a 1000 distinct genetic markers [3-6] . Individual variations in body weight, however, are not constant throughout the life course but are the cumulative effect of a complex interplay between genetics, environment and behavior. As such, investigating the temporal aspect of the genetic basis of obesity and its interaction with the changing obesogenic (obesity-promoting) environment is fundamental for gaining a full understanding of the mechanisms that underpin this complex trait.

Although several studies have investigated the age-dependent genetic effects of BMI loci, the majority of them have been performed in cross-sectional analyses [7-10] or in longitudinal designs with narrow periods of the life course [11-14], primarily childhood and early adulthood. In addition, the age-dependent effects of these loci have not been studied in the context of generational differences because the generational and age related effects on obesity cannot be properly disentangled with cross-sectional data. Therefore, which specific genes influence weight development in middle and late adulthood, two crucial life stages for health and disease, and whether these effects are different among different generations, remains poorly understood.

To investigate how genetic variation affects individual BW development during adulthood, we performed a GWAS on the bodyweight trajectories of 4619 subjects from the Doetinchem cohort that have been followed for a period of 25 years. Specifically, association analyses were performed on the imputed BW at 5 different ages (30, 40,50,60,70 years) and on average BW and overall BW gain/loss. To derive the imputed longitudinal measures from the heterogeneous set of BW measurements, a linear mixed-effects model (LMM) was constructed that accounted for height, gender, year of birth, age and individual deviations in average BW, BW gain/loss, and BW fluctuations. To correct for age and generation effects on BW in an unbiased manner, the age-generation-gender interaction was modelled semi-parametrically using a cubic bivariate penalized (P-) spline function, where the penalty parameter was estimated as an inverse variance together with the other model parameters in the REML procedure (Suppl. Methods).

We identified rs663129 (*MC4R*) in the average BW GWAS as the only locus that reached genome-wide significance (rs663129-A: beta=1.3, *p*=1.1E-09), while no loci in the GWAS on the other traits reached genome-wide significance (Fig. 1 and Suppl. Fig. S1). Given that our study was underpowered for a GWAS on BW or obesity related traits, we subsequently focussed our analysis on the 1192 BMI associated variants that had been previously discovered by Locke *et al.* (2015) [3]. From this set we found that the minor allele of rs2075650 (*TOMM40/APOE*) was associated with weight loss (rs2075650-G: beta=-3.0, *p*=4.9E-07) and that rs543874 (*SEC16B*) showed a transient effect on BW at the age of 50 (rs543874-G: beta=1.3, p=8.8E-06). Next, we took forward the loci of rs663129, rs2075650 and rs543874 for a more detailed analysis of their longitudinal effect on BW, by including the interaction between the allele dosage and a P-spline function of age in the LMM. As the GWAS results implied, rs663129 has a stable effect on BW during lifetime, whereas rs2075650 and rs543874 showed age-dependent BW profiles (Fig. 2).

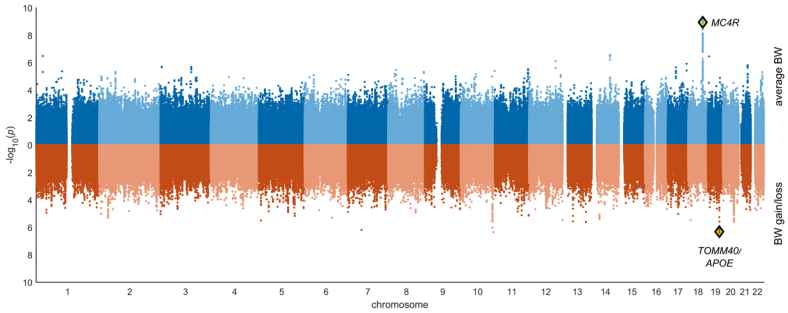

Figure 1. Miami plot showing the association results of the GWAS on average BW (blue) and BW gain/loss (red). Diamonds indicate associations that are significant in the set of known BMI loci as reported by Locke et al. (2015) [3].

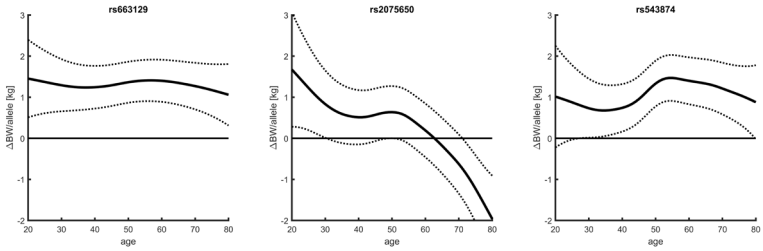

Figure 2. Allele-dependent BW trajectories for rs663129-A (*MC4R*), rs2075650-G (*TOMM40/APOE*) and rs543874-G (*SEC16B*).

Interestingly, the *TOMM40/APOE* locus almost reached genome-wide significance in our study with 4619 subjects, whereas it just surpassed this threshold in the study of Locke *et al.* (2015) [3], which was done in 339K individuals. Clearly, therefore, performing GWAS cross-sectionally may fail to pick up genetic modifiers with transient but negligible effects on average BW, thus demonstrating the added value of a longitudinal design. This is also reflected by the estimated heritability of this locus, which is higher when accounting for the longitudinal effect (h^2 = 0.063%) as compared to the cross-sectional (constant) heritability (h^2 = 0.011%) (Suppl. Tab. S1). Inspecting the BW-age profile of the rs2075650-G allele in detail, we estimated a positive effect on BW of 0.64 +/- 0.33 kg/allele at the age of 50, after which a period of consistent weight loss set in arriving at -1.96 +/- 0.55 kg/allele at the age of 80.

Rs2075650 is a very pleiotropic locus that, apart from BMI, has strong associations with plasma cholesterol [15], Alzheimer's disease (AD) [16] and all-cause mortality [17]. Because of linkage disequilibrium rs2075650 is generally considered to be a proxy for the *APOE* e4 variant (rs429358), although conditional analyses on BMI have shown that rs2075650 also has a small independent contribution [18]. *APOE* e4 is well-known as the strongest genetic risk factor of late onset AD, suggesting that the progressive weight loss that we observed after the age of 50 could be part of the cascade of pathological processes underlying AD. In fact, even though AD is diagnosed at least 10 years later on the average popuplation[19] , the prospective study of [20] on familial (early onset) AD has shown that markers for AD are affected up to 15 years prior to diagnosis.

The *SEC16B* locus has been identified already more than a decade ago by Speliotes *et al.* (2010) [21], but its age-dependent effect on weight development is novel. Interestingly, phenome-wide analysis of rs543874 using the summary statistics collected in GWAS ATLAS [22] showed that it also modulates the age of menarche [23] (Suppl. Fig. S5). Specifically, the minor allele rs543874-G is associated with a weight increase at the age of 55 of 1.47 +/- 0.30 kg/allele (p = 2.6E-07) in our study and is associated with

a decrease in age at menarche of 0.069 years/allele in the study of Day *et al.* (2017) [23]. In addition, Locke *et al.* (2015) [3] found that rs543874 has a sex-specific effect on BMI and is higher in women, which we confirmed with the sex-stratified analysis of the allele specific BW trajectories (Suppl. Fig. S4). Although it remains currently unknown by which mechanisms *SEC16B* acts on both obesity and the age of puberty in girls, these findings make it plausible that the distinct longitudinal pattern of the BW trajectory of rs543874 is somehow linked to age-related hormone changes.

In conclusion, we here report the first GWAS performed on longitudinal changes in BW during adult life. Even though our study was underpowered for weight and obesity related traits, we identified two known obesity loci with pronounced age-dependent effects on BW. Indeed, variance decomposition showed that BW is remarkably stable during adult life, explaining 87.3% of the individual variance, as compared to 4.7% and 8.0% of weight gain and residuals, respectively (Suppl. Fig. S9). Consequently, GWAS on longitudinal effects in weight development need to be extremely well powered in order to detect loci that are significant at the genome-wide scale, which is a clear weakness of our study. Nevertheless, we were still able to detect the association of the *TOMM40/APOE* locus with overall BW gain because of the relatively wide age range of our cohort at baseline (20-59 years) and the long duration of our study (25 years).

The novelty of our approach for longitudinal GWAS is to combine the traditional random intercept, random slope design with a semi-parametric approach for modelling continuous-time functions that can accommodate nonlinear dependencies of the main covariates. As a result, we were able to estimate the age-dependent effect of rs2075650 and rs543874 on BW development in an unbiased way, which provides important clues for their involvement in biological processes such as cognitive decline and age-related hormone changes. In addition, via this approach, we confirmed the increased exposure of the obesogenic environment on the younger generations in our cohort, but we also uncovered a subtle change in the mean BW development in our cohort from the year 2000 onwards that might be linked to a general change in life style (Suppl. Fig. S10). We believe that by including other relevant environmental and lifestyle factors such as physical activity levels and smoking status we will be able to further dissect BW variation and thereby gain insight into this complex trait.

ACKNOWLEDGMENTS

The authors would like to thank the field workers of the Municipal Health Services in Doetinchem for their contribution to the data collection, P. Vissink for logistic management and A. Blokstra for data management (all from the National Institute for Public Health and the Environment).

REFERENCES

1. Seidell, J. C. & Halberstadt, J. The global burden of obesity and the challenges of prevention. *Ann. Nutr. Metab.* 2015; 66 Suppl 2: 7–12.
2. Dixon, J. B. The effect of obesity on health outcomes. *Mol. Cell. Endocrinol.* 2010; 316:104–108.
3. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015; 518: 197–206.
4. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature.* 2015; 5181:87–196.
5. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* 2017; 49:1458–1467.
6. Huang, T. *et al.* Genetic predisposition to central obesity and risk of type 2 diabetes: Two independent cohort studies. *Diabetes Care.* 2015; 38:1306–1311.
7. Hertel, J. K. *et al.* FTO, type 2 diabetes, and weight gain throughout adult life: a meta-analysis of 41,504 subjects from the Scandinavian HUNT, MDC, and MPP studies. *Diabetes.* 2011; 60: 1637–1644.
8. Graff, M. *et al.* Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Hum. Mol. Genet.* 2013; 22:3597–3607.
9. Brandkvist, M. *et al.* Separating the genetics of childhood and adult obesity: a validation study of genetic scores for body mass index in adolescence and adulthood in the HUNT Study. *Hum. Mol. Genet.* 2021; 29:3966–3973.
10. Winkler, T. W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet.* 2015; 11:1–42.
11. Choh, A. C. *et al.* Gene-by-age effects on BMI from birth to adulthood: The fels longitudinal study. *Obesity.* 2014; 22: 875–881.
12. Graff, M. *et al.* BMI loci and longitudinal BMI from adolescence to young adulthood in an ethnically diverse cohort. *Int. J. Obes.* 2017; 41:759–768.
13. Hardy, R. *et al.* Life course variations in the associations between FTO and MC4R gene variants and body size. *Hum. Mol. Genet.* 2009; 19: 545–552.
14. Buscot, M. J. *et al.* Longitudinal association of a body mass index (BMI) genetic risk score with growth and BMI changes across the life course: The Cardiovascular Risk in Young Finns Study. *Int. J. Obes.* 2020; 44:1733–1742.
15. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 2013; 45:1274–1285.
16. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* 2022; 54:412–436.
17. Deelen, J. et al. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell.* 2011; 10:686–698.
18. Kulminski, A. M. et al. Independent associations of TOMM40 and APOE variants with body mass index. *Aging Cell.* 2019; 18: e12869.
19. Koedam EL, Lauffer V, van der Vlies AE, van der Flier WM, Scheltens P, Pijnenburg YA. Early-versus late-onset Alzheimer's disease: more than age alone. *J Alzheimers Dis.* 2010; 19:1401-1408.

20. Bateman, R. J. et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N. Engl. J. Med.* 2012; 367:795–804.
21. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* 2010; 42:937–948.
22. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* 2019; 51:1339–1348.
23. Day, F. R. et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* 2017; 49:834–841.

SUPPLEMENTARY METHODS

Population

The Doetinchem Cohort Study (DCS) is a prospective population-based cohort study including 7769 men and women aged 20-59 years at baseline living in Doetinchem, predominantly from Dutch Caucasian descent, between in 1987-1991 (round 1). Participants of the first round were invited for follow-up examinations every 5 years (Suppl. Fig. S6). The response rates varied between 75% and 80% in all rounds. The study design of DCS has previously been described in more detail [1,2]. All participants gave written informed consent and the study was approved according to the guidelines of the Helsinki Declaration by the external Medical Ethics Committee of the University Medical Center Utrecht (UMCU).

Measurements

Demographic characteristics, medical history of chronic diseases, medication use, and lifestyle factors were collected using standardized questionnaires. Bodyweight and height were measured by trained staff in each round with participants wearing light indoor clothing without shoes, with emptied pockets. Bodyweight was measured to the nearest 100 g on calibrated scales.

Genotyping, quality control and imputation

Genomic DNA was isolated from venous blood samples of 5088 individuals at the Dutch National Institute for Public Health and the Environment, and genotyped in the HUman GEnomics Facility (HUGE-F) Rotterdam using the Illumina Infinium Global Screening Array-24 Kit (GSA) (Illumina Inc., San Diego, California, United States of America) [3]. The R package GenABEL 1.8-0 [4], was used to perform the quality control for both participants and genetic variants.

Participants were excluded if : 1) there was a sex mismatch (n=45), 2) the sample call rate was <95% (n=20), 3) heterozygosity rate was high (false discovery rate (FDR) <1%) (n=37), 4) samples were duplicates (n=18) or monozygotic twins (n=1) (one individual per pair) and 5) participants were widely diverged (i.e. being genetically distant based on visual inspection of a genomic Principal Component (gPC) plot) regarding their genetic background based on the first two gPCs that were constructed using a kinship matrix, in two steps (n=114). First, the more distant participants compared to the group as a whole were excluded. Next, as a single iterative step new gPCs were generated in the remaining sample population and additional participants were removed.

Genetic variants were excluded when: 1) minor allele frequencies (MAF) were < 1/(2*5088), i.e. the chance of finding the allele once in the study population, thus representing monomorphic variants (n=109129), 2) genotype call rates were <98% (n=8005), 3) variants were not in Hardy-Weinberg equilibrium (FDR<0.2) (n=0), and 4) X-linked markers were likely to be autosomal (n=421). Subsequently, the HRC-1000G-check-bim.pl script from Rayner [5] was used for quality control and to convert the Plink genotype data to separate VCF files per chromosome. This pre-imputation step of quality control filtered additional SNPs based on genotype call rate < 98% (n=15013) and Hardy-Weinberg p<10-6 (n=0). Finally, genotypes were imputed to the Haplotype Reference Consortium (HRC) panel (version r1.1 2016) [6] with the Michigan Imputation Server [7] using genomic build NCBI 37. Pre-phasing was performed on the imputation server with Eagle v2.3 [6] and imputation with Minimac3 [8]. After quality control and imputation of the GSA-data and matching with subjects with more than two bodyweight measurements, 4619 participants were left for further analyses.

Statistical Analyses

General. Analysis of longitudinal bodyweight data was performed using a linear mixed-effects model (LMM) that takes into account height, age, gender, year of birth (YOB), genetic variation and individual weight fluctuations. The LMM is described in detail in the next section. All analyses were performed on bodyweight instead of BMI because dividing BW by height squared does not resolve the association with height and also causes the data to become strongly heteroscedastic, which is undesirable from a statistical modelling point of view (Suppl. Fig. S11). Descriptive analyses were performed in SPSS Statistics 23 and LMM computations and visualizations were done in Matlab. LMM analyses were performed with the restricted maximum likelihood (REML) approach and statistical inference was done using the Satterthwaite approximation for the denominator degrees of freedom. Analyses were adjusted for multiple testing using Bonferroni's method and a (corrected) p-value < 0.05 was considered statistically significant.

Genetic association analyses. GWAS was performed on the individual estimates of the random intercept and random slope in the LMM (Eq. 1) without the genetic component. For the GWAS on the BW slope, the BW intercept was included as covariate and vice versa. Next, the fitted LMM was used to impute individual BW measures at ages 30, 40, 50, 60 and 70, and GWAS was performed on these values. Imputed values were set to missing if the corresponding age was more than 5 years before the youngest age at which that subject was screened, or when it was more than 5 years after the oldest age at which that subject was screened. GWAS was performed using rvtests; analyses were corrected for population stratification by including the kinship matrix and the first 4 principal genetic components. Association results were considered both for genome-wide

significance and for statistical significance of a subset of known BMI loci as reported by Locke *et al.* (2015) [9].

For loci that were identified in the first stage the age-dependent effect on bodyweight was determined by refitting the LMM after including the interaction term of the B-spline function of age and the genotype of the specific locus. This analysis was performed separately for each locus, first by pooling men and women together and then by including an interaction term for sex and the age-dependent allele effect.

Variance decomposition and heritability analysis. Explained variance of the model's components was determined in the fitted LMM without genetic components (Suppl. Fig. S9). Specifically, variations were computed for the components of interest and were then squared and summed. The variance in the age-generation interaction was further investigated by computing the estimated means for men and women for a range of different years of birth, and plotting the resulting trajectories as a function of age and time/occasion of measurement (Suppl. Fig. S10). Heritability of the identified loci was calculated as the ratio of the sum of squared variation of the age-dependent allele effect and the total sum of squares of BW variation. In addition, the heritability of the static (age-independent) allele effect was computed by fitting the LMM without the interaction between the allele and the B-spline function of age.

Phenome-wide analysis. The degree to which the loci with distinct age-dependent effects on bodyweight development were pleiotropic was investigated by looking up the summary statistics of these loci in the GWA studies that were collected in the GWAS ATLAS [10] . Subsequently, associations with traits for which $p < 10^{-10}$ were plotted (Suppl. Fig. S5).

References

1. Verschuren WMM, Blokstra A, Picavet HSJ, Smit HA. Cohort profile: the Doetinchem Cohort Study. *Int J Epidemiol*. 2008;37:1236–1241.
2. Picavet HSJ, Blokstra A, Spijkerman AMW, Verschuren WMM.Cohort Profile Update: The Doetinchem Cohort Study 1987-2017: Lifestyle, Health and Chronic Diseases in a Life Course and Ageing Perspective. *Int J Epidemiol*. 2017;46:1751-1751.
3. Illumina Infinium Global Screening Array-24 Kit. Illumina Inc., San Diego, California, United States of America; Available from: http://www.glimdna.org/assets/2017-infinium-global-screening-array-illumina-data-sheet.pdf.
4. Karssen LC, van Duijn CM, Aulchenko YS. The GenABEL Project for statistical genomics. F1000Res. 2016;5:914. Epub 2016/06/28.
5. Rayner. HRC-1000G-check-bim.pl script. Available from: http://www.well.ox.ac.uk/~wrayner/tools.
6. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279-83. Epub 2016/08/23.
7. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284-7. Epub 2016/08/30.
8. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44:955-9. Epub 2012/07/24.
9. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518:197–206.
10. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet*. 2019; 51:1339–1348.

A linear mixed-effects model of bodyweight change over time

Model definition. We chose to model the individual variation in bodyweight in terms of a set of subject-specific random effects of continuous time functions. In order to properly set up the random effects of the LMM, we first performed an exploratory analysis to investigate the dependency of average bodyweight on age by fitting a penalized B-spline function on the whole dataset (Suppl. Fig. S7, left). The resulting dependency exhibited a distinct nonlinear shape, resembling that of a negative exponential function or root function. Subsequently we fitted both types of functions, resp. $BW(t) = \beta_0 + \beta_1 e^{-\frac{t}{\tau}}$ and $BW(t) = \beta_0 + \beta_1 t^{\alpha}$, and optimized the parameters in the exponent (Suppl. Fig. S7, right). The negative exponential function with $\tau = 29.1$ years described the data slightly better than the optimized power law function, and fitted the data much better than the linear slope model (power law function with $\alpha = 1$). Consequently, we decided to model the main age dependency at the individual and population level as a negative exponential function with $\tau = 29.1$ years.

To account for all major sources of variation in the bodyweight data – height, gender, age, year of birth (YOB), genetic variation and individual variation – we constructed an LMM containing components representing each of these sources of variation:

$$BW_{ij} = BW_{ij}^{\text{height}} + BW_{ij}^{\text{age}} + BW_{ij}^{\text{gene}} + BW_{ij}^{\text{subj}} + \varepsilon \tag{1}$$

In (1), BW_{ij} is the bodyweight of subject i measured at occasion j, BW_{ij}^{height} the gender dependent effect of height and squared height on BW, BW_{ij}^{age} the mean BW per age, corrected for gender and generation, BW_{ij}^{gene} the genetic principal components and the age-dependent allele effect of the tested SNP, BW_{ij}^{subj} the subject-specific BW variations with age, and ε the random fluctuations in BW that are independent between measurements and subjects.

In detail, the four bodyweight variation components are defined as follows.

$$BW_{ij}^{\text{height}} = h_i(1-s_i)\beta_{\text{male},1}^{\text{height}} + h_i^2(1-s_i)\beta_{\text{male},2}^{\text{height}} + h_i s_i \beta_{\text{female},1}^{\text{height}} + h_i^2 s_i \beta_{\text{female},2}^{\text{height}} \tag{2}$$

with h_i the height of subject i at baseline, $s_i = 0$ if subject i is male and $s_i = 1$ if subject i is female, and β_*^{height} the fixed effect parameters.[1]

The population mean BW per age is defined as

$$\begin{aligned} BW_{ij}^{\text{age}} \quad = \quad & (1-s_i)\exp\left(-\frac{age_{ij}}{\tau}\right)\beta_{\text{male}}^{\text{age}} + (1-s_i)\,\mathbf{B}_{ij}^{\text{YOB-time}}\mathbf{b}_{\text{male}}^{\text{YOB-time}} \\ & + s_i \exp\left(-\frac{age_{ij}}{\tau}\right)\beta_{\text{female}}^{\text{age}} + s_i\,\mathbf{B}_{ij}^{\text{YOB-time}}\mathbf{b}_{\text{female}}^{\text{YOB-time}} \end{aligned} \tag{3}$$

with age_{ij} the age of subject i at occasion j, τ the exponential coefficient discussed previously, $\mathbf{B}_{ij}^{\text{YOB-time}}$ the bivariate cubic B-spline design matrix of the YOB-time interaction,[2] and $\mathbf{b}_*^{\text{YOB-time}}$ the vector of spline coefficients penalized by the integral of the second order derivative of the spline basis functions $\mathbf{D}^{\text{YOB-time}}$ times a smoothing parameter $\lambda^{\text{YOB-time}}$. Importantly, the penalized B-splines in (3) can be rewritten in terms of a mixed effects model, where the smoothing parameter

[1] Throughout this section we will use β for fixed effect parameters, u for zero mean random effect parameters that are identically and independently distributed, and b for random effect parameters with a nonzero mean. Moreover, boldface notation will be used for vectors and matrices while italics will be used to denote scalar variables.

[2] We preferred here to model the YOB-time interaction instead of the YOB-age interaction (where *age* equals *time* minus *YOB*), since the longitudinal BW measurements cover a rectangular area in the former notation and a parallelogram in the latter (Suppl. Fig. S6), which suffers more from boundary effects when fitting a bivariate B-spline function.

corresponds to the ratio of σ_ε^2 with the variance of the random effects.[3] The first term in (3) models the coarse dependency of bodyweight on age where τ controls the curvature of the BW-age dependency, while the second term models the deviations from the exponential model and the additional generation effect.

The genetic component is defined as

$$BW_{ij}^{\text{gene}} = \mathbf{C}\,\boldsymbol{\beta}^{\text{genetic PCs}} + g_i \mathbf{B}_{ij}^{\text{age}} \mathbf{b}^{\text{age-gene}} \tag{4}$$

with $\mathbf{C}\,\boldsymbol{\beta}^{\text{genetic PCs}}$ the genetic principal components and coefficients that correct for population stratification, g_i the number of effect alleles of subject i, $\mathbf{B}_{ij}^{\text{age}}$ the cubic B-spline design matrix of age, and $\mathbf{b}^{\text{age-gene}}$ the vector of spline coefficients penalized by the integral of the second order derivative $\mathbf{D}^{\text{age-gene}}$ of the spline basis functions times a smoothing parameter $\lambda^{\text{age-gene}}$.

Finally, the individual BW term is defined as

$$BW_{ij}^{\text{subj}} = u_i^{\text{base}} + f_{ij}^{\text{inc}} u_i^{\text{inc}} + \mathbf{B}_{ij}^{\text{age-var}} \mathbf{u}_i^{\text{age-var}} \tag{5}$$

with u_i^{base} the zero-mean random intercept, u_i^{inc} the zero-mean random BW increase, where the BW increase function f_{ij}^{inc} is defined as the non-negative exponential function in (3) with zero mean for subject i, $\mathbf{B}_{ij}^{\text{age-var}}$ the cubic B-spline design matrix of the slow BW fluctuations with age, and $\mathbf{u}_i^{\text{age-var}}$ the corresponding vector of random effects. An important difference between the use of the B-spline function to model age in (5) with respect to (3) and (4) is that in (5) the spline function contains a small number of knots and is penalized by the identity matrix, whereas in (3) and (4) a much higher number of knots can be chosen. As a result, in the latter case the higher order derivatives need to be penalized, resulting thus in a rank deficient penalization matrix $\mathbf{D}$.

An illustration of an individual BW trajectory and the corresponding mixed model components is shown in Suppl. Fig. S8.

Computational considerations. The mixed model can be rewritten by combining the fixed and random effects such that the latter have zero mean and are identically and independently distributed

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_0\mathbf{u}_0 + \mathbf{Z}_1\mathbf{u}_1 + \boldsymbol{\varepsilon} \tag{6}$$

where $\mathbf{y}$ is the vector that contains all BW measurements, $\mathbf{X}\boldsymbol{\beta}$ the fixed effect design matrix and parameters, which consists of (2) and the non-penalized part of (3) and (4), $\mathbf{Z}_0\mathbf{u}_0$ the random effect design matrix and parameters of the population mean, which consists of the penalized part of (3) and (4), and $\mathbf{Z}_1\mathbf{u}_1$ the individual, subject-specific random effect design matrix and parameters, which consists of (5). For large datasets it is important to distinguish between $\mathbf{Z}_0$ and $\mathbf{Z}_1$, because $\mathbf{Z}_0$ has a fixed number of columns and remains dense when the number of subjects increases, whereas the number of columns in $\mathbf{Z}_1$ will grow with the number of subjects and will become more sparse. Specifically, $\mathbf{Z}_1$ has a block structure that needs to be properly exploited in order to calculate the inverse of the covariance matrix efficiently for large studies and, consequently, to make the likelihood maximization procedure computationally feasible. We developed a MATLAB package that is optimized for fitting mixed models of type (6) for large studies.

[3] In general, for a regression formula $\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$ with the coefficients $\mathbf{b}$ penalized by a factor $\lambda\mathbf{b}^{\mathrm{T}}\mathbf{D}\mathbf{b}$ (called Tikhonov or $L2$ regression), $\mathbf{b}$ can be decomposed into a set of fixed and random effects via eigendecomposition of $\mathbf{D}$. Let matrices $\mathbf{N}$ and $\mathbf{V}$ be defined as $\mathbf{V}^{\mathrm{T}}\mathbf{D}\mathbf{V} = \mathbf{I}$, $\mathbf{D}\mathbf{N} = \mathbf{0}$, $\mathbf{V}^{\mathrm{T}}\mathbf{N} = \mathbf{0}$ and $\mathbf{N}^{\mathrm{T}}\mathbf{N} = \mathbf{I}$, then it can be shown that $\mathbf{b} \sim \mathcal{N}(\mathbf{N}\boldsymbol{\beta}, \lambda^{-1}\sigma_\varepsilon^2\mathbf{V}\mathbf{V}^{\mathrm{T}})$ and consequently that the regression formula can be rewritten as a mixed model $\mathbf{y} = (\mathbf{X}\mathbf{N})\boldsymbol{\beta} + (\mathbf{X}\mathbf{V})\mathbf{u} + \boldsymbol{\varepsilon}$ with fixed effects $\boldsymbol{\beta}$ and random effects $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\sigma_\varepsilon^2\mathbf{I})$.

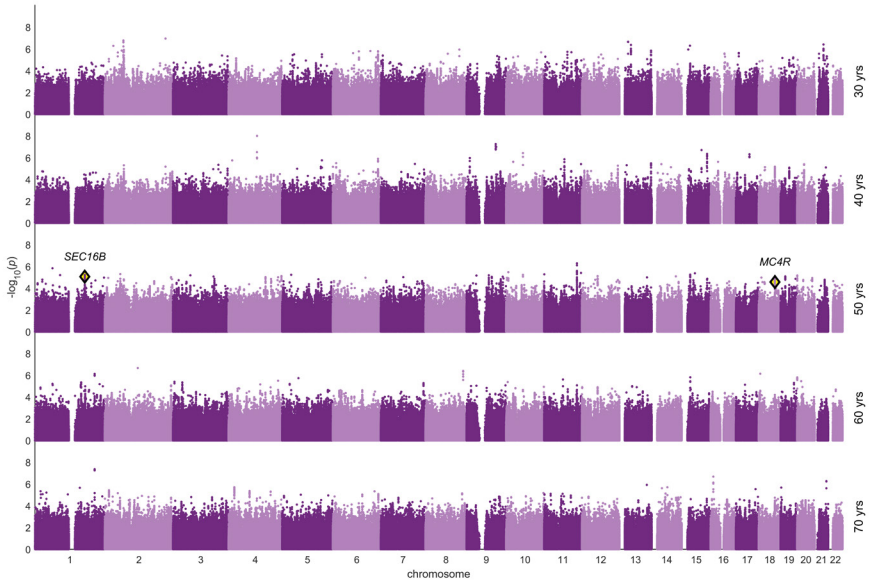

Figure S1: Manhattan plots showing GWAS results on imputed bodyweight at ages 30, 40, 50, 60 and 70 years. The linear mixed-effects model without genetic effects was fitted on the BW trajectories of all subjects and subsequently used to predict the individual deviations from the mean at the aforementioned ages. GWA analysis was then performed on these imputed age-specific residual measures of bodyweight. Results were considered for genome-wide significance and for statistical significance of a subset of known BMI loci (Locke *et al.*, 2015); significant loci according to the second criterion are visualized as diamonds.

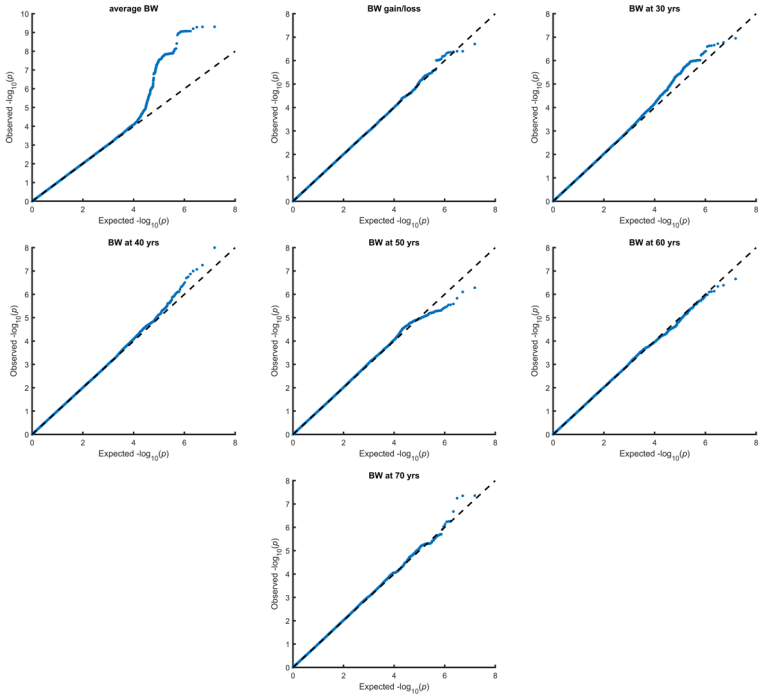

Figure S2: QQ plots of the association statistics of each of the GWA studies performed here: average BW, BW gain/loss, and BW imputed at 5 different ages.

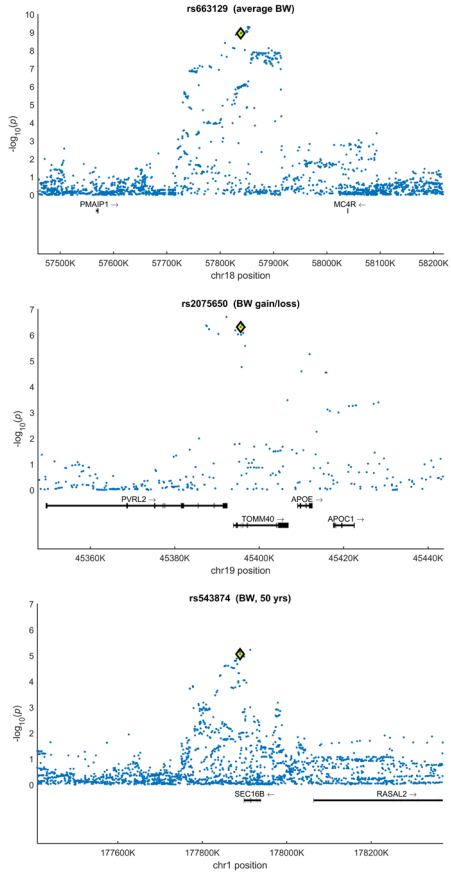

Figure S3: Regional association plots of the identified loci and the corresponding bodyweight related trait.

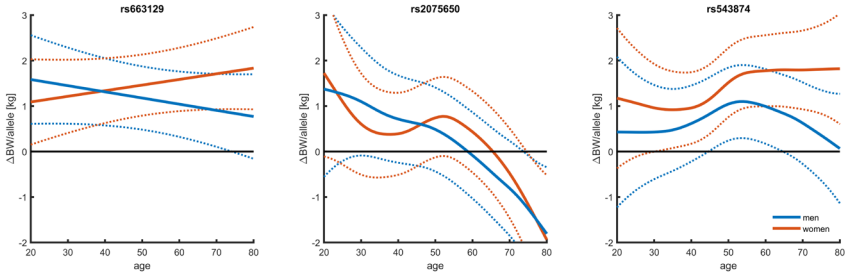

Figure S4: Allele-dependent bodyweight trajectories for rs663129 (*MC4R*), rs2075650 (*TOMM40/APOE*) and rs543874 (*SEC16B*), stratified for sex. No significant differences are found in the trajectories of rs663129 and rs2075650 between men and women, while the trajectories of rs543874 start to diverge after midlife and reach statistical significance ($p < 0.05$) after the age of 73.

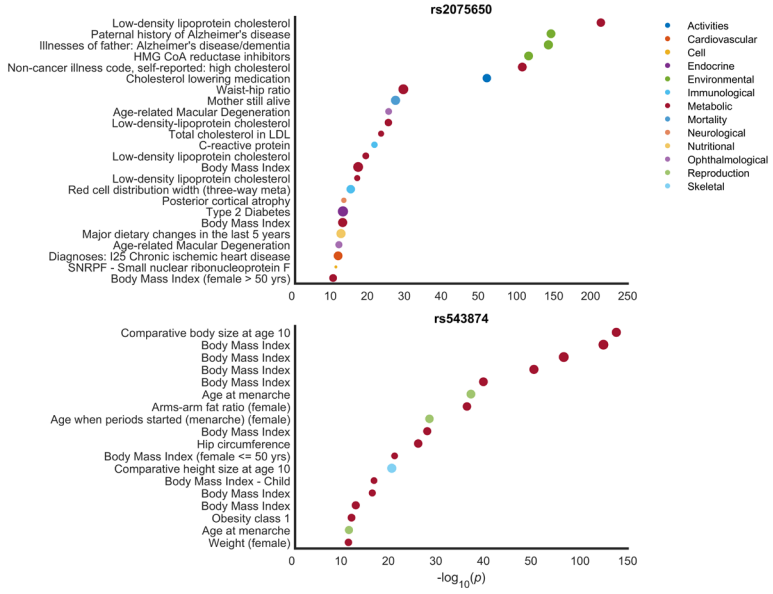

Figure S5: Phenome-wide analysis of rs2075650 (*TOMM40/APOE*) and rs543874 (*SEC16B*). Associations with other traits ($p < 10^{-10}$) are shown of both loci as based on the summary statistics collected in the GWAS ATLAS (https://atlas.ctglab.nl/PheWAS; Watanabe *et al.* (2019)). In case a locus was associated with multiple related traits that were analyzed in the same study only the most significant trait is shown in the graph. The dot size shows the relative cohort size of the corresponding GWAS.

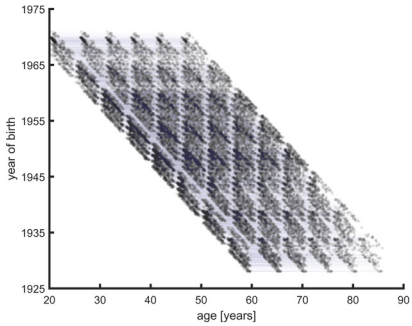

Figure S6: Bodyweight measurements (dots) in the Doetinchem Cohort Study started in the period 1987–1991 and subjects were invited for follow-up examinations every 5 years, for a total of 6 rounds.

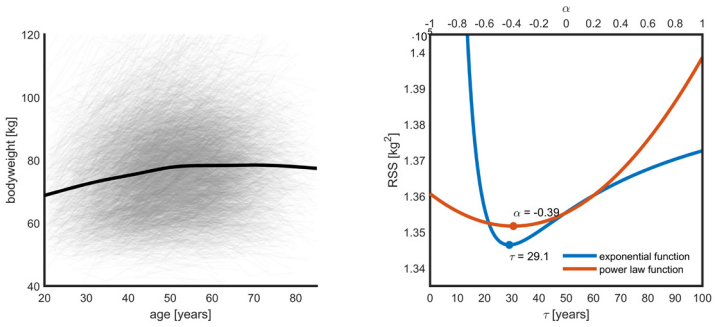

Figure S7: Exploratory analysis of the longitudinal patterns in the bodyweight data. Average bodyweight increases with age in a nonlinear fashion (left), suggesting that a mixed-effects model with random linear slopes does not accurately describe the data. Fitting a negative exponential ($BW(t) = \beta_0 + \beta_1 e^{-\frac{t}{\tau}}$) and power-law function ($BW(t) = \beta_0 + \beta_1 t^{\alpha}$) on the data, the negative exponential function best describes the nonlinear dependency of average bodyweight on age (right). RSS: residual sum of squares

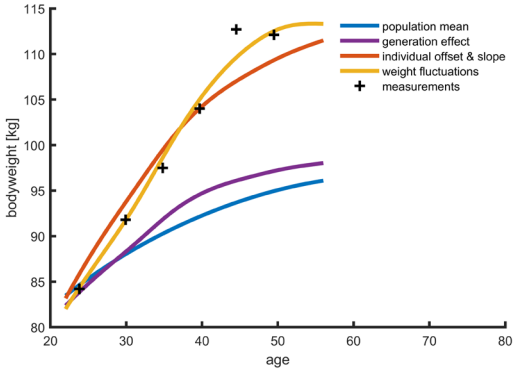

Figure S8: Illustration of the bodyweight trajectory from a male subject and the corresponding LMM components without genetic effects. The global population mean consists of a negative exponential function (blue) and the additional effect of the YOB-age interaction (purple). The individual variations consists of a negative exponential function with subject-specific intercept and negative exponential slope (red) and the personal BW fluctuations with age (yellow). The YOB-age interaction is modelled by a penalized bivariate cubic B-splines function, and the personal BW fluctuations with a cubic B-spline function with a low number of knots. Since this is an individual from the younger generation, the generation effect is higher than the population mean, which means that individuals from this birthyear are expected to be heavier during midlife than individuals from the older generations. YOB: year of birth.

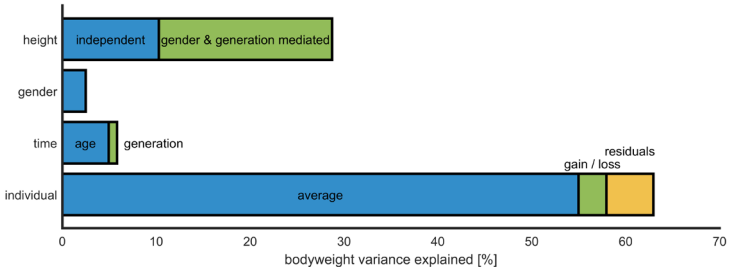

Figure S9: Variance decomposition of bodyweight. The explained variance of the LMM components was determined for the model without genetic effects. As expected, height proved to be the covariate that explained the largest proportion of BW variation, corresponding to 18.4% due to gender and generation associated height differences and 10.3% independent height association. Gender explained 2.5% independently of height, age explained 5.0%, the year of birth explained 0.9% independently of height, and the remaining individual variation amounted to 63.0%. Interestingly, of the individual variation the vast majority (87.3%) was explained by the random intercept in the LMM, which corresponds to the age-averaged deviation from the population mean. The random slope explained only 4.7% and the residuals 8.0%. LMM: linear mixed-effects model.

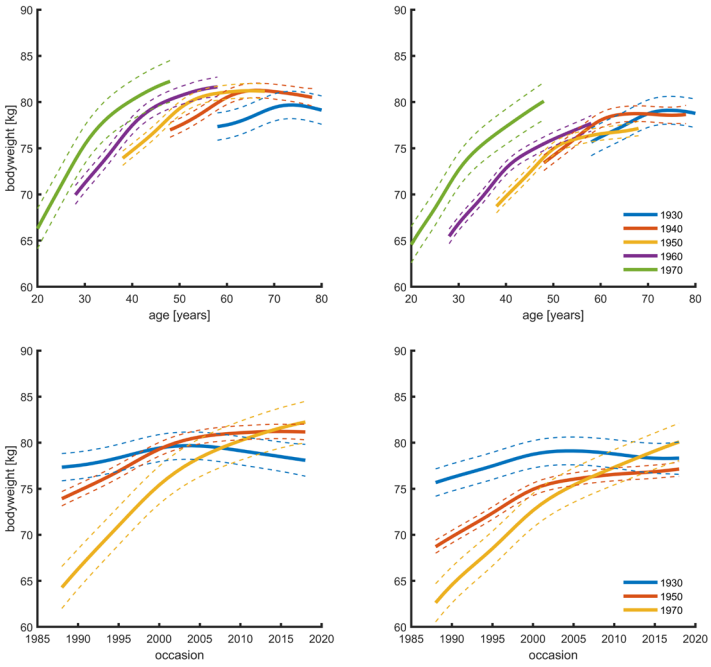

Figure S10: Investigation of the mean bodyweight trajectories for men (top left) and women (top right) for different generations confirmed our earlier findings that younger generations were progressively heavier that the older ones, except for women of the oldest birth cohort (Hulsegge *et al.*, 2014). Interestingly, the continuous-time bodyweight estimates provided an indication of a purely chronological effect on bodyweight development in men and women that occurred roughly half-way during our study. Plotting the bodyweight dependency as a function of the time of measurement for men (bottom left) and women (bottom right) and stratified by birth cohort, showed a clear turning point in bodyweight development in our study population between 2000 and 2005. Although the origin of this change in curvature is unknown, we hypothesize that it could be related to the increased public awareness of the health consequences of obesity.

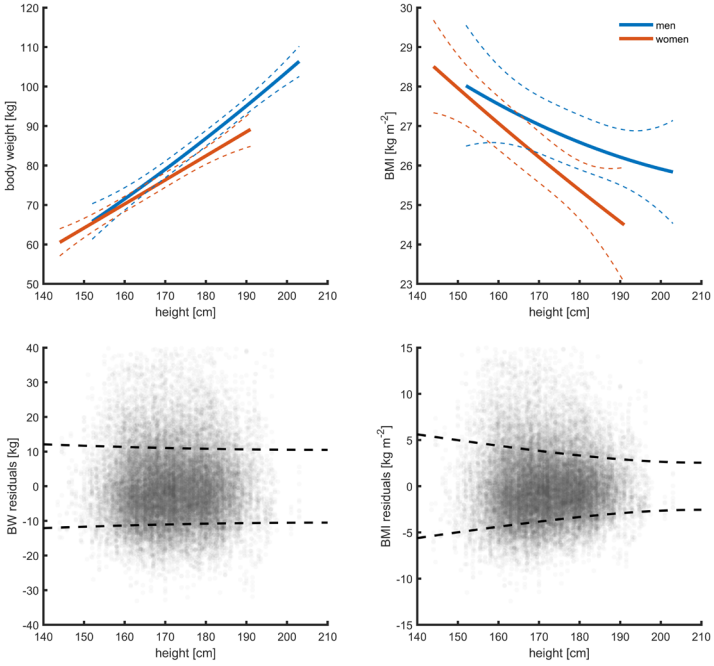

Figure S11: Bodyweight dependency on height (top left) and BMI dependency on height (top right) for men and women. Whereas BW is strongly dependent on height, dividing by squared height attenuates this dependency but does not eliminate it. The residual variance in bodyweight shows no dependency on height (bottom left) whereas BMI has pronounced heteroscedasticity (bottom right). BW has similar variances for short and tall individuals (variance of 137 and 111 kg^2 for individuals with a height of resp. 150 and 200 cm), whereas the variance in BMI of short individuals is almost four times that of tall individuals (variance of 24.9 and 7.0 kg^2m^{-4} for individuals with a height of resp. 150 and 200 cm). Height-dependent variance was estimated on the LMM residuals e through maximum likelihood, assuming a model $\mathbf{e} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{H}\sigma_h^2 + \mathbf{I}\sigma_e^2\right)$, with $\mathbf{H}$ a diagonal matrix with elements $h_{ij} = (3 - \text{height [cm]}/70)^2$.

locus	h^2 [%] constant	h^2 [%] age-dependent
rs663129	0.357	0.361
rs543874	0.221	0.246
rs2075650	0.011	0.063

Table S1: Constant and age-dependent heritability estimates of the identified bodyweight loci. The effect of rs663129 on bodyweight is almost constant, which is confirmed by the fact that the constant and age-dependent heritability estimates are approximately the same. The rs543874 locus has a small time-varying effect on bodyweight as can be seen in Fig. 2, but the largest effect on bodyweight is constant as is shown by the small difference between the heritability estimates. In contrast, the effect of rs2075650 is almost entirely age-dependent given that the age-dependent heritability is almost 6 times larger than the constant heritability.

References

Hulsegge, G., Picavet, H., Blokstra, A., Nooyens, A., Spijkerman, A., van der Schouw, Y., Smit, H., and Verschuren, W. (2014). Today's adult generations are less healthy than their predecessors: generation shifts in metabolic risk factors: the doetinchem cohort study. *Eur J Prev Cardiol.*, **21**(9), 1134–1144.

Locke, A., Kahali, B., Berndt, S., Justice, A., Pers, T., Day, F., Powell, C., *et al.* (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**(7538), 197–206.

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T., van der Sluis, S., Andreassen, O., Neale, B., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet.*, **51**(9), 1339–1348.