Supporting Information

Ho et al. 10.1073/pnas.0910878107

SI Materials and Methods

Identification of the Risk Allele. The HapMap sample (1) has genotyped, in several populations, far more SNPs than are genotyped in our sample. Using the HapMap as a template, we found a proxy in our sample for these SNPs of interest. Fig. S1 shows the LD (quantified using the r^2 statistic) using the HapMap data from a European sample (CEU) from HapMap data Phase III/Release 2, Feb 2009, NCBI b36 assembly. There are two SNPs in high LD with the associated risk marker, rs8050136 (r^2 = 0.936 to rs1421085; r^2 = 1 to rs17817449) and rs3751812 (r^2 = 0.958 to rs1421085; r^2 = 0.979 to rs17817449). Both of these tagging SNPs were linked to both verified risk SNPs and to each other, so we chose to use only one SNP, rs3751812, with the highest mean correlation (mean r^2 = 0.9875) to both verified risk SNPs. The initially identified *FTO* variant (rs9939609) (2) is also in high LD with the chosen tagging SNP (r^2 = 0.979; Fig. S1). At the tagging SNP, rs3751812, the risk allele is the T allele and the frequency in the population of those sampled was 16.02% (T/T), 46.12% (T/G), and 37.86% (G/G). Subjects with one or two risk alleles were pooled into a single "carriers" group due to small sample size of subjects carrying two risk alleles (n = 33).

DNA Isolation and SNP Genotyping Methods. DNA was isolated from B lymphocyte cells taken from blood (3) and extracted (4) using standard procedures. Seven milliliters of EDTA blood was extracted using the QIAamp DNA Blood Maxi Kit (Qiagen). Samples were processed according to the manufacturer's protocol. Genomic DNA samples were analyzed on the Human610-Quad BeadChip (Illumina) according to the manufacturer's protocols (Infinium HD Assay; Super Protocol Guide; Rev. A, May 2008). Before the initiation of the assay, 50 ng of genomic DNA from each sample was examined qualitatively on a 1% Tris-acetate-EDTA agarose gel for visual signs of degradation. Any degraded DNA samples were excluded from further analysis. Samples were quantitated in triplicate with PicoGreen reagent (Invitrogen) and diluted to 50 ng/μl in Tris-EDTA buffer (10 mM Tris, 1 mM EDTA, pH 8.0). Two hundred nanograms of DNA was then denatured, neutralized, and amplified for 22 h at 37 °C (termed the MSA1 plate). The MSA1 plate was then fragmented with FMS reagent (Illumina) at 37 °C for 1 h and then precipitated with 2-propanol and incubated at 4 °C for 30 min. The resulting blue precipitate was then resuspended in RA1 reagent (Illumina) at 48 °C for 1 h. The samples were then denatured (95 °C for 20 min) and immediately hybridized onto BeadChips at 48 °C for 20 h. BeadChips were then washed and subjected to single-base extension and staining. Finally, the BeadChips were coated with XC4 reagent (Illumina), dessicated, and imaged on the BeadArray Reader (Illumina).

Correlation Between BMI and Years of Education. We wanted to investigate whether any association between BMI, the *FTO* risk allele, and brain structure was mediated by some other directly measurable variable, such as educational level. In a separate analysis of healthy elderly individuals from the Cardiovascular Health Study (5), BMI was found to be weakly, but significantly correlated to the number of years of education completed (n = 256; r^2 = 0.024; P = 0.012). In the sample of 206 healthy subjects in ADNI, there was no relation between BMI and educational level, measured in years (n = 206; $r^2 = 4.8 \times 10^{-4}$; P = 0.76). This result may be because there is more variation in the educational level for the Cardiovascular Health Study sample (σ^2 = 17.91) than the ADNI sample (σ^2 = 8.70); even so, educational level did not explain the effects of genotype detected here, nor was it associated with *FTO* genotype.

White Matter Burden and Brain Structure. White matter hyperintensity (WMH) was assessed in the space of the MDT on the basis of the signal intensities of coregistered T_1-, T_2-, and PD-weighted scans and on the basis of population statistics regarding the spatial distribution and neighborhood structure of white matter lesions throughout the brain (6). The method provides white matter hyperintensity measures that agree strongly with FLAIR-based gold-standard measures. White matter burden (WMB) was defined as $\log_{10}$ of the WMH volume in cubic centimeters. The log transform was used so the values fit better a normal distribution (Jarque–Bera test; P = 0.3294).

Cumulative Distribution Function (CDF) Plots. CDF plots, based on the above correlations, are shown in Fig. S3. These plots can be used to visually rank the degree of statistical association between brain structure and any external parameter, such as genetic variation, BMI, or WMB. CDF plots that rise more sharply at the origin indicate variables that correlate more strongly with brain structure. Structural brain deficits were more strongly correlated with BMI (green curve) than with WMB (red curve) or the *FTO* tagging SNP, rs3751812 (blue curve). The black line represents a reference; the point where the black line intersects the other curves represents the statistical threshold that can be applied to the map to control the false discovery rate at 5%; in other words, when thresholded in this way, the statistical map is expected to contain only 5% false positive suprathreshold voxels. CDFs can be used in this way to rank the effects of covariates (7) where higher CDFs tend to indicate stronger effect sizes, on the basis of the definition that a more lenient FDR q-controlling critical P value reflects an improvement in the overall fit of the covariate in the map. Even so, this must be qualified because the maps from which the CDFs were created relate to entirely different brain regions and networks, as well as having stronger or weaker associations within those regions, so the highest CDF plot cannot be interpreted as finding the most powerful associations. Rather, it shows the proportion of P values at multiple statistical thresholds and presents a convenient way to compare the total extents of maps at multiple statistical thresholds (albeit conflating different regions).

Description of Statistical Models. Several different statistical models were used throughout the manuscript; these are explained in the next section.

Association Between BMI and *FTO* Genotype. The association between BMI and *FTO* genotype was conducted using a standard multiple regression model, of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \qquad [S1]$$

where y is a vector whose components represent BMI for each subject, x_1 is a binary vector representing presence or absence of the risk allele of the *FTO* gene at the tagging SNP, x_2 is a vector representing the age of each subject, x_3 is a vector representing the sex of each subject, and ε is an error term. An ordinary least-squares cost function was used to estimate each β_i parameter, and the significance and value of the relevant β_i parameter are reported in the main text.

Obesity-Associated Risk Alleles and Brain Structure. Associations between *FTO* genotype and brain structure were examined using

exactly the same multiple regression model as in Eq. **S1**, except that the dependent variable y was a vector representing the determinant of the Jacobian matrix of deformation at a particular voxel (which measures the volume of that region relative to the mean template). The regression model was fitted at each voxel in the brain to determine the association of this risk allele to structure across the entire brain, after adjusting for any effects of age and sex. Fig. 1 (*Upper*) in the main text shows the P value of the β_1 parameter (expressing the *FTO* effect) in regions that survived the false discovery rate correction for multiple comparisons, and Fig. 1 (*Lower*) shows the value of β_1 in areas with a significant P value.

BMI and Brain Structure. Associations between BMI and brain structure were again examined using a similar multiple regression model to Eq. **S1**, except that the dependent variable, y, was a vector representing the determinant of the Jacobian matrix of deformation at a particular voxel and the independent measure, x_1, was a vector representing the BMI across subjects. The regression was conducted at each voxel in the brain to determine the association of BMI to structural differences, in the subject sample, across the entire brain. Fig. 2 (*Upper*) in the main text shows the P value of the β_1 parameter (expressing the BMI effect) in regions that survived the false discovery rate correction for multiple comparisons; Fig. 2 (*Lower*) shows the value of β_1 in areas with a significant P value.

White Matter Burden and Brain Structure. Associations between white matter hyperintensity and brain structure were again evaluated with a similar statistical model to Eq. **S1**, except that the dependent variable, y, was a vector representing the determinant of the Jacobian matrix of deformation at a particular voxel and the independent measure, x_1, was a vector representing the $\log_{10}$ transform of the white matter hyperintensity volume (cm^3) across subjects. The WMB measure is a single number for each subject, but its effect on brain volumes was assessed voxelwise. The regression was run at each voxel in the brain to determine the association of WMB to structure across the entire brain. Fig. S2 (*Upper*) shows the P value of the β_1 parameter (expressing the WMB effect) in regions that survived the false discovery rate correction for multiple comparisons; Fig. S2 (*Lower*) shows the value of β_1 in areas with a significant P value.

Glucose by Risk Allele Interactions on Brain Structure. We also assessed whether there were any interactions between (*i*) glucose levels by (*ii*) the presence or absence of the *FTO* risk allele. The interactions assessed the effects of these two variables on brain structure, after controlling for age, sex, glucose levels, and risk allele effects according to the regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_4 + \varepsilon, \qquad \textbf{[S2]}$$

where y is a vector representing the determinant of the Jacobian matrix of deformation at a particular voxel (a measure of regional brain volume), x_1 is a binary vector representing presence or absence of the risk allele of the *FTO* gene at the tagging SNP, x_2 is a vector representing the age of each subject, x_3 is a vector representing the sex of each subject, x_4 is a vector representing glucose levels across subjects, and x_1x_4 represents a term that will fit if there is a significant interaction between glucose levels and risk allele status. This regression was conducted at each voxel across the brain, and it was found that the β_5 parameter (the interaction term) was not significantly different from zero in the brain when using an FDR correction for multiple comparisons.

WMB by Risk Allele Interactions on Brain Structure. The statistical model used for this analysis was exactly the same as in Eq. **S2** with the exception that x_4 represents a vector representing WMB levels across subjects, and x_1x_4 represents the interaction between WMB levels and risk allele status. Similarly, this regression was conducted at each voxel across the brain, and it was found that the β_5 parameter was not significantly different from zero in the brain when using an FDR correction for multiple comparisons.

Interpretation of Regression Coefficients and Volume Units. The regression coefficients reported in the main text show the size of the effect on brain structure (Figs. 1 and 2 and Figs. S2, S4, and S5). There are two equivalent representations of the regression coefficients. The percentages in the main text do represent the group differences in mean volume, but all of the group differences are denominated in units of volume *relative to* an unbiased standard brain template—a template specially constructed to be balanced, or unbiased, by including equal numbers of people from each genotype group (Fig. 1 and Table S2). This percentage is not exactly the same as the percentage volume difference between carrier and noncarrier groups if one of those groups is used as a reference. For example, if the carriers were 10% smaller than the template and the noncarriers were 10% bigger, we report this as a 20% difference, although it could be considered a 22.2% difference ($100 \times 0.2/0.9$), if the carriers were used as a reference, or an 18.2% difference ($100 \times 0.2/1.1$) if the noncarriers were used as a reference. If one group were ill and the other healthy, we would tend to express deficits relative to the healthy group. But as all subjects here are healthy, we used the population mean as a reference to define the units of percentage for volume, so that would lead to the noncarriers being considered as having a volume excess. As such, the percentage units are quantified as a percentage of the size of structures in the mean template, rather than relative to the size of structures in the noncarrier group, but they do reflect differences between the two groups.

To avoid confusion, the color bar and figure legends in the main text specify units of ΔTemplate Volume (mm^3). This unit is convertible to a percentage difference if the population mean volume is considered as the reference. Each voxel has a volume of 1 mm^3, and percentage volume reduction or excess relative to the template may be written in terms of the change in the template volume (i.e., a 5% reduction relative to the template would indicate a deficit of 0.05 mm^3 at that voxel). Similarly the regression coefficients for continuous variables such as BMI and WMB (Fig. 2 and Figs. S2 and S5) represent the difference—as a proportion of the mean template volume—per unit of the continuous variable. The units of this regression slope are mm^3 per $kg{\cdot}m^{-2}$, in the case of BMI, for example. However, neither representation is exactly the same as the percentage volume difference between the carrier and noncarrier groups, if one of those groups is chosen as a reference. If the volume differences are expressed relative to the noncarrier group (who have larger volumes), the percentage differences would be marginally smaller (as the reference value is higher). We used the unbiased template as a reference, rather than the noncarriers, as arguably all subjects are normal, so the noncarriers might be considered as having a higher volume than the healthy population mean and the carriers a lower volume.

Image Correction Across Sites. Rigorous guidelines for standardizing MRI scans across study site and across time are detailed in ref. 8. Briefly, images were calibrated with phantom-based geometric corrections to ensure consistency across scanners. Additional image corrections included (*i*) correction of geometric distortions due to gradient nonlinearity, (*ii*) adjustment for image intensity inhomogeneity due to B1 field nonuniformity using calibration scans, (*iii*) reducing residual intensity homogeneity, and (*iv*) geometric scaling according to a phantom scan acquired for each subject to adjust for scanner- and session-specific calibration errors. Each incoming image file was quality checked for

medical abnormalities and image quality. Because of the extent of these corrections and that a large number of sites (58) acquired data, study center was not included as a covariate in these analyses. The scans used in this study were collected from 53 of the 58 sites.

In addition, the ADNI scanning protocol was developed after a rigorous preparatory phase in which we and others made sure that the scan volumes were reproducible and stable across repeated scanning, using the same volume quantification method as in this paper (9). Even so, we did check if any one scanner vendor was overrepresented in the *FTO* allelic groups. The table below represents the breakdown of subjects by scanner vendor:

Breakdown of scanners used by vendor

	FTO risk allele group (n = 128)	*FTO* nonrisk allele group (n = 78)
GE Healthcare	62	38
Phillips Medical Systems	16	14
Siemens Medical Solutions	50	26

As this was a multisite study, rigorous calibration and cross-site standardizations were performed to ensure the comparability and stability of the imaging measures; we checked if any one scanner vendor was overrepresented in the *FTO* allelic groups, and no differences were found in the scanner vendor used (χ^2_2 = 1.42; P = 0.4916).

1. Frazer KA, et al.; International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861.
2. Frayling TM, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889–894.
3. Neitzel H (1986) A routine method for the establishment of permanent growing lymphoblastoid cell lines. *Hum Genet* 73:320–326.
4. Lahiri DK, Bye S, Nurnberger JI, Jr, Hodes ME, Crisp M (1992) A non-organic and non-enzymatic extraction method gives higher yields of genomic DNA from whole-blood samples than do nine other methods tested. *J Biochem Biophys Methods* 25:193–205.
5. Lopez OL, et al. (2003) Prevalence and classification of mild cognitive impairment in the Cardiovascular Health Study Cognition Study: Part 1. *Arch Neurol* 60:1385–1389.
6. Schwarz CG, Fletcher D, DeCarli C, Carmichael OT (2009) Fully-automated white matter hyperintensity detection with anatomical prior knowledge and without FLAIR. *Inf Process Med Imaging* 21:239–251.
7. Hua X, et al. (2009) Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: An ADNI study of 515 subjects. *Neuroimage* 48:668–681.
8. Jack CR, Jr, et al. (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* 27:685–691.
9. Leow AD, et al. (2006) Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *Neuroimage* 31:627–640.

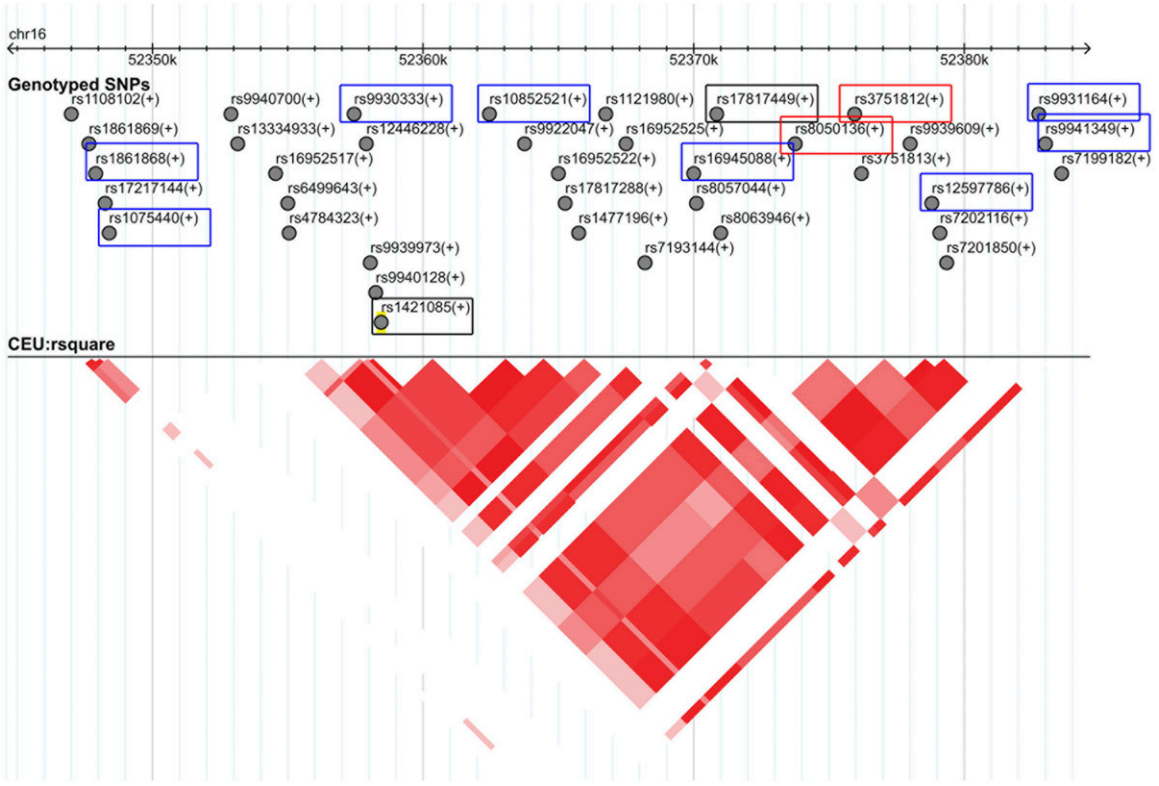

Fig. S1. Linkage disequilibrium (LD) patterns from the HapMap European (CEU) sample near the previously verified SNPs associated with obesity. (*Upper*) The SNPs genotyped in the HapMap sample are shown as gray circles representing their location on the 16th chromosome (see scale, at top), whereas the gray circles are lined up with this scale to indicate their position. The SNPs associated with obesity are shown in black boxes. The SNPs genotyped both in the HapMap and in our sample are shown in blue and red boxes. The two red boxes show those SNPs in our sample with the highest LD to the SNPs associated with obesity. (*Lower*) The LD pattern (as an r^2 statistic) between every pair of SNPs shown. Darker red represents a higher correlation. This image is modified from the HapMap website (http://www.hapmap.org/).

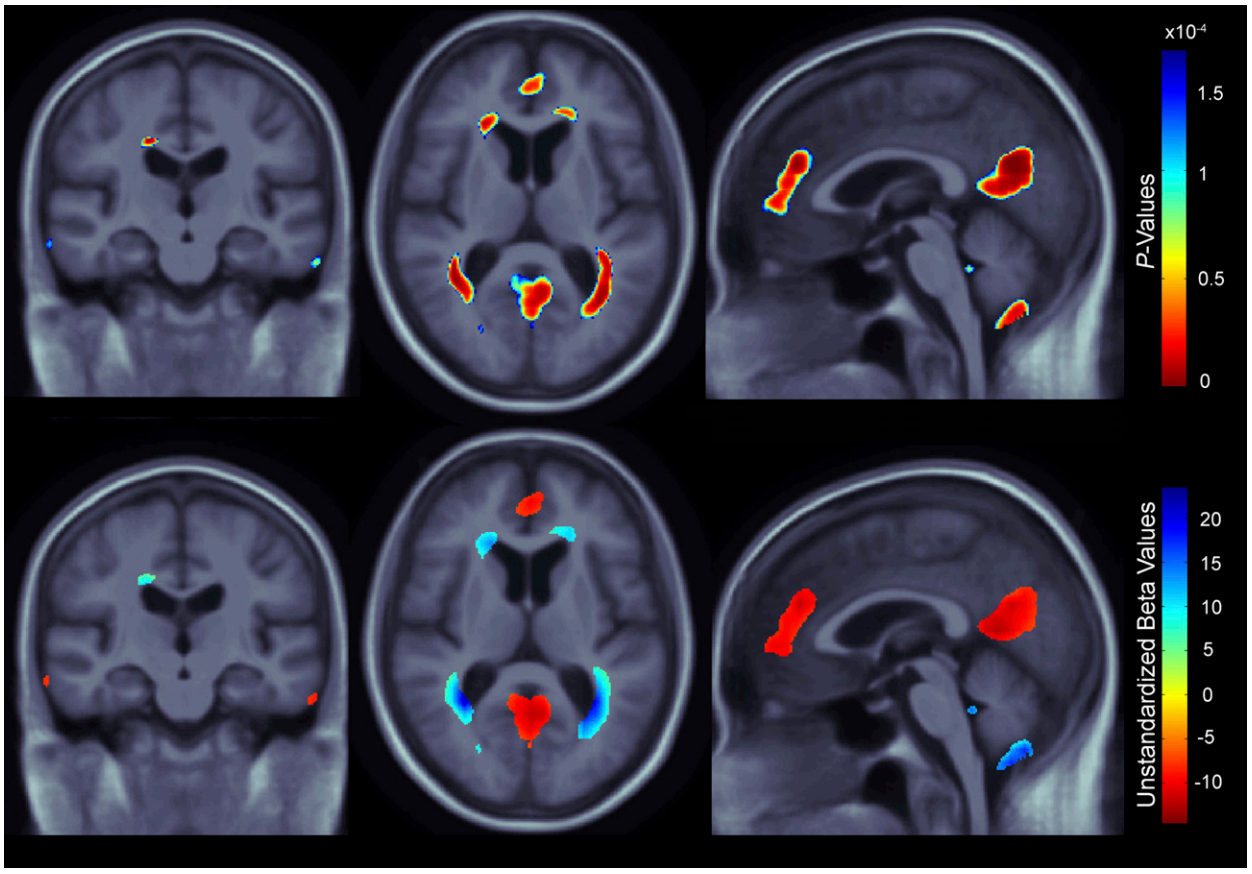

Fig. S2. (*Upper*) 3D maps show areas where regional brain volumes were significantly associated with log-transformed whole brain white matter hyperintensity (WMH) volume (cubic centimeters) in healthy elderly subjects (n = 169). The $\log_{10}$ of whole brain WMH volume is referred to as white matter burden (WMB). (*Lower*) In the significant areas, the regression coefficients (unstandardized beta values) are shown at each voxel. These represent the estimated degree of tissue excess or deficit at each voxel, for each log cubic centimeter increase in whole brain WMH volume, after statistically controlling for effects of age and sex on brain structure. Areas that are positively associated with WMB are shown in blue (tissue expansion mainly represented as CSF expansion found at the ventricle boundaries) and areas that are negatively associated with WMB are shown in red (volume deficits) (*Lower*). The majority of areas that are statistically significant are associated with a *deficit* in brain volume in those with greater WMB, which is as expected. All maps were significant after standard correction for multiple comparisons using FDR (q = 0.05; critical uncorrected P = 0.0016). Images are displayed in radiological convention (left side of the brain shown on the right) and are displayed over a study-specific brain image template (MDT).

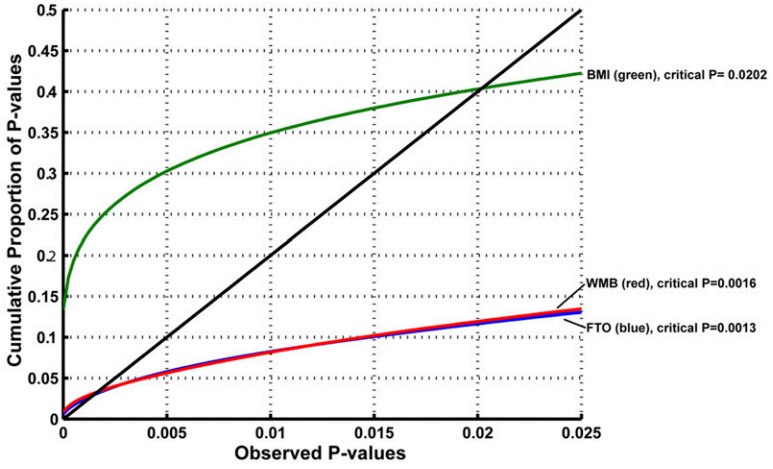

Fig. S3. The cumulative distribution of P values observed for the maps of association created above. Brain structure is more strongly associated with BMI (*green line*; critical uncorrected P = 0.0202) than with the *FTO* tagging SNP, WMB (*red line*; critical uncorrected P = 0.0016), or rs3751812 (*blue line*; critical uncorrected P = 0.0013). Even so, both associations are significant, according to the FDR criterion. The black line represents a reference: The point where it intersects the green or blue lines, other than at the origin, represents a statistical threshold that can be applied to the map to control the false discovery rate at q = 0.05 (i.e., 5% of the suprathreshold voxels are expected to be false positives).

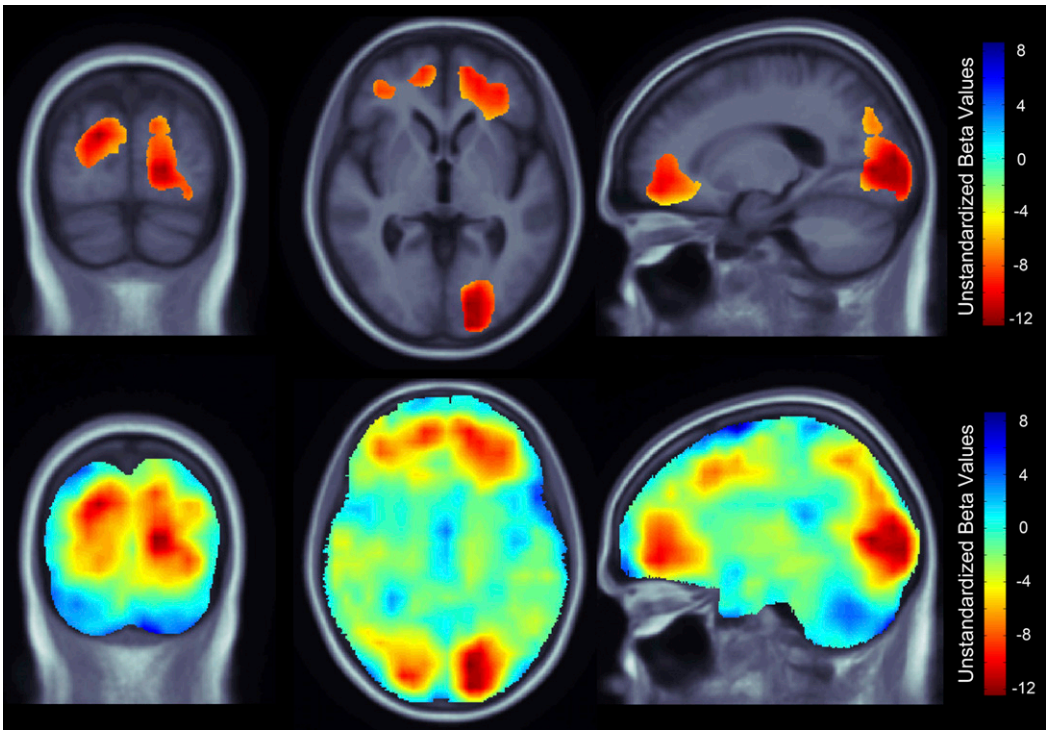

Fig. S4. Regression coefficient (unstandardized beta) maps for regional brain volume correlating with *FTO*, after adjusting for age and sex (*Upper*; same as Fig. 1 *Upper*) compared to unthresholded regression coefficient maps for regional brain volume correlating with *FTO*, after adjusting for age, sex, and BMI (*Lower*). Unthresholded maps are shown (*Lower*) as there were no significant regions after correcting at 5% FDR in the whole brain analysis after adjusting for age, sex, and BMI. This is perhaps to be expected as *FTO* influences BMI. Images are displayed in radiological convention (left side of the brain shown on the right) and are displayed over a study-specific brain image template (MDT).

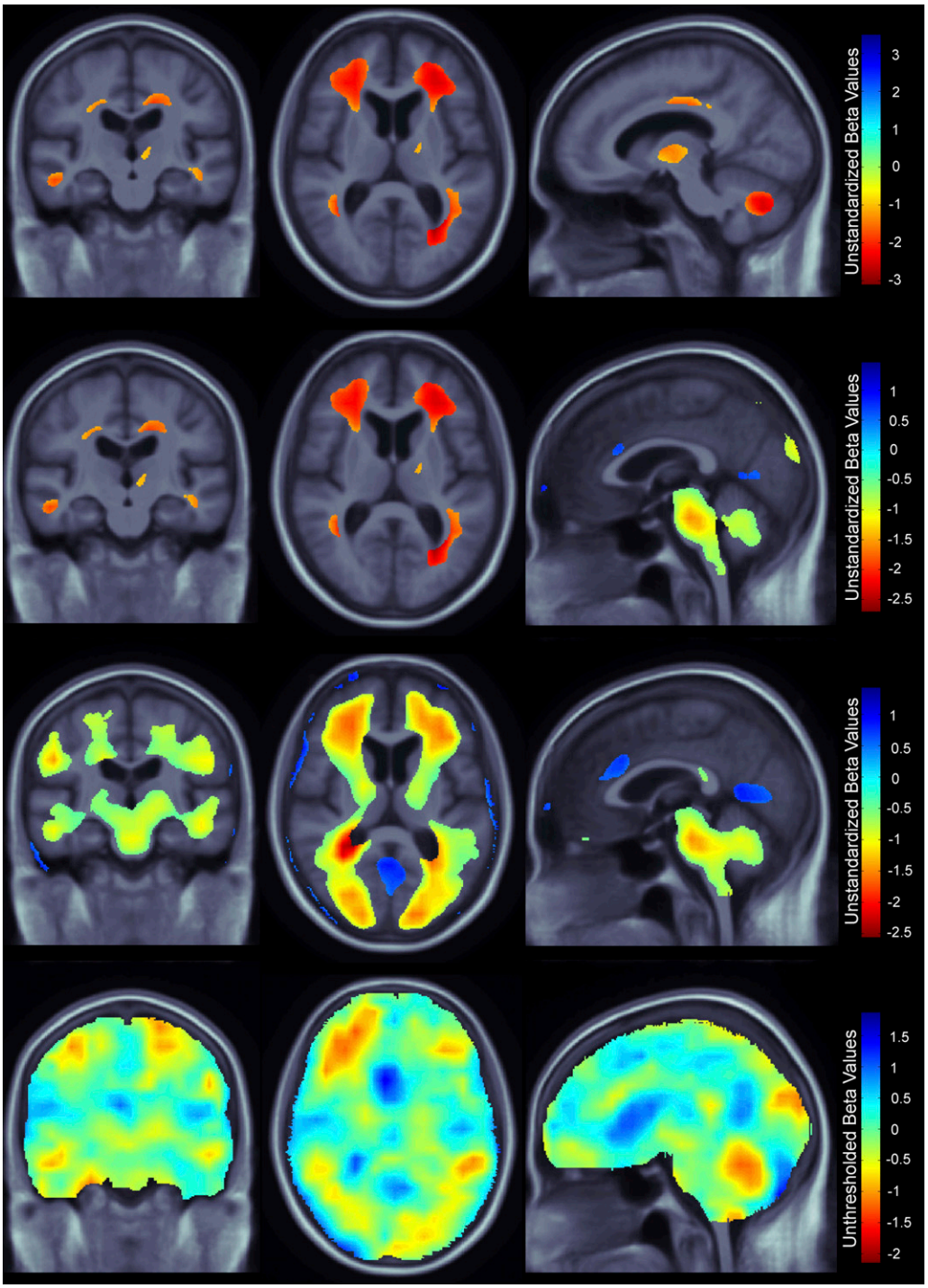

Fig. S5. 3D maps show areas where regional brain volumes were significantly associated with BMI in subjects carrying two copies of the risk allele (n = 33; critical uncorrected P = 0.0022) (first row), subjects carrying a single copy of the risk allele (n = 95; critical uncorrected P = 0.0113) (second row), and subjects carrying at least one copy of the risk allele (n = 128, critical uncorrected P = 0.016) (third row). Intriguingly, volumes were not statistically associated with BMI, after FDR correction, in *FTO noncarriers* (fourth row, unthresholded maps). The regression coefficients (unstandardized beta values) are shown at each voxel. These coefficients represent the estimated degree of tissue excess or deficit at each voxel, as a percentage, for each unit gain in body mass index, after statistically controlling for effects of age and sex, on brain structure. Images are displayed in radiological convention (left side of the brain shown on the right) and are displayed over a study-specific brain image template (MDT).

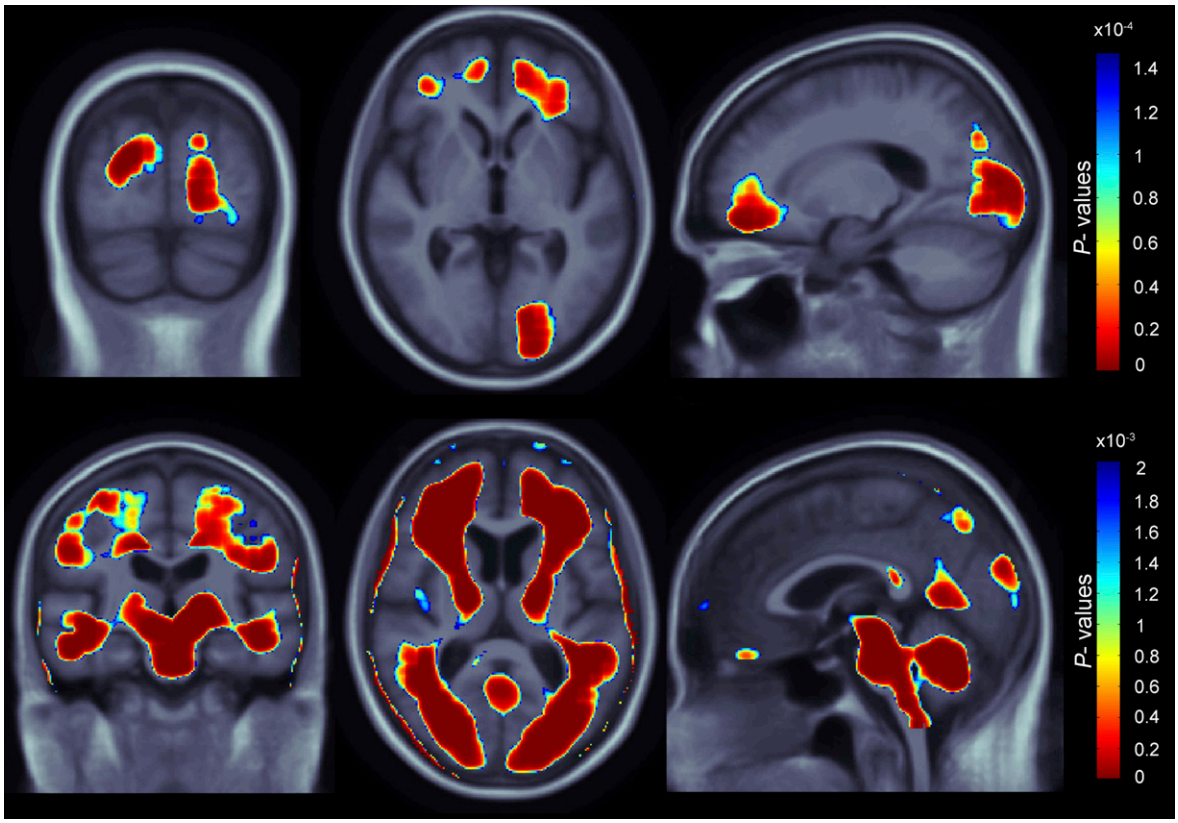

Fig. S6. 3D maps show areas where regional brain volumes were significantly associated with having one or more risk alleles of the *FTO* gene ($n = 206$; critical uncorrected $P = 0.0015$) (*Upper*) and BMI ($n = 206$; critical uncorrected $P = 0.021$) (*Lower*) based on a $\log_{10}$ transformation of the Jacobian determinants. These maps are consistent with the significance maps based on the Jacobian determinants (Figs. 1 and 2).

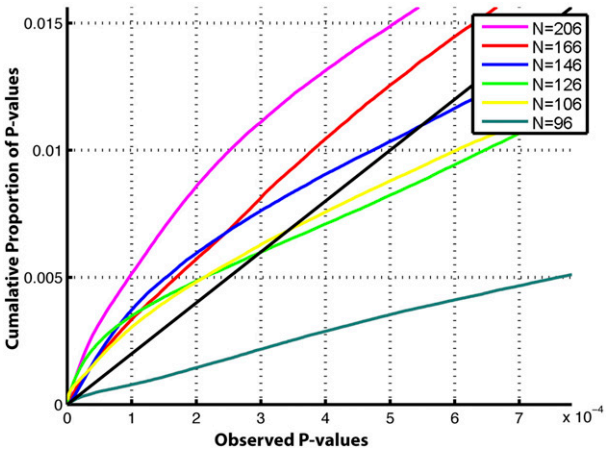

Fig. S7. Effects of varying the sample size. Plots of the cumulative proportion of *P* values are shown for correlations between volume differences at each voxel in the whole brain and carrying the risk allele of the *FTO* gene as sample size (n) is reduced. In general, smaller sizes show lower critical *P* values than those computed from large sample sizes. The approximate minimal sample size needed to repeat the finding by randomly removing subjects from our sample was calculated to be 96.

Table S1. Demographic information for the 206 subjects

Genotype groups	Nonrisk homozygous genotype (G/G)	Risk heterozygous genotype (T/G)	Risk homozygous genotype (T/T)	Statistical Results
Sample size (*n*)	78 (44 M/ 34 F)	95 (47 M/ 48 F)	33 (20 M/ 13 F)	$\chi^2_2 = 1.54$; $P = 0.46$
Age (years)	76.1 ± 4.7	76.4 ± 4.8	75.6 ± 6.1	$F_{2,203} = 0.284$; $P = 0.75$
Body mass index (BMI)	25.6 ± 4.1	27.0 ± 4.7	27.2 ± 4.1	$F_{2,203} = 2.66$; $P = 0.072$
Education	16.3 ± 2.5	16.6 ± 2.8	14.7 ± 2.9	$F_{2,203} = 5.85$; **$P = 0.0034$**
Glucose (mg/dL)	58.0 ± 9.71 (*n* = 32)	59.6 ± 9.19 (*n* = 49)	63.9 ± 7.68 (*n* = 15)	$F_{2,93} = 2.12$; $P = 0.13$
Cholesterol (mg/dL)	195.0 ± 36.5 (*n* = 77)	196.3 ± 38.2 (*n* = 93)	193.6 ± 49.1 (*n* = 31)	$F_{2,198} = 0.0609$; $P = 0.94$
White matter hyperintensity volume (cm^3)	2.4 ± 1.8 (*n* = 69)	2.9 ± 3.2 (*n* = 77)	2.7 ± 2.0 (*n* = 23)	$F_{2,166} = 0.653$; $P = 0.52$
History of hypertension	(37/78) 47.4%	(44/95) 46.3%	(10/33) 30.3%	$\chi^2_2 = 1.30$; $P = 0.52$
History of stroke	(2/78) 2.6%	(0/95) 0%	(0/33) 0%	$\chi^2_2 = 3.23$; $P = 0.20$
History of cardiovascular disease	(53/78) 67.9%	(66/95) 69.5%	(16/33) 48.5%	$\chi^2_2 = 1.16$; $P = 0.56$
Mini mental state examination	29.3 ± 0.874	29.1 ± 0.934	29.0 ± 0.937	$^*\chi^2_2 = 1.61$; $P = 0.45$
BMI Groups (kg/m^2)	BMI < 25	25 ≤ BMI < 30	BMI ≥ 30	
Global clinical dementia rating	0	0	0	NA
Mini mental state examination	29.4 ± 0.8	29.0 ± 0.9	29.1 ± 1.0	$^*\chi^2_2 = 7.43$; **$P = 0.0244$**

The mean ± SD is shown for each measure; clinical measures are categorized by number of risk alleles at the obesity-associated tagging SNP and body mass index groups. ANOVAs, performed on each variable, show whether the mean clinical measure differed significantly across groups. χ^2 tests were performed for categorical variables, where subscripts indicate degrees of freedom. A nonparametric Kruskal–Wallis one-way ANOVA was used to assess measures that were not normally distributed (e.g., MMSE scores). White matter hyperintensity values were $\log_{10}$ transformed before statistical tests. F, females; M, males; NA, not analyzed. Boldface type indicates statistical significance ($P < 0.05$).

Table S2. Demographic information for all 40 subjects included in the construction of the minimal deformation template (MDT)

MDT Subject demographics	
Sample size (*n*)	40
Nonrisk genotype (G/G)	20
Risk genotype (T/G)	10
Risk genotype (T/T)	10
Age (years)	77.0 ± 4.6
Sex (male)	20
Body Mass Index (BMI)	26.2 ± 3.9

The MDT was designed to be an unbiased representation of the sample. Equal numbers of subjects with risk and nonrisk genotypes were included, and equal numbers of men and women were included in the MDT group. There was no statistical difference in the age of the subjects used in the MDT relative to the other subjects in the sample (76.95 ± 4.56 years vs. 75.95 ± 5.08 years; $t_{204} = 1.143$; $P = 0.2542$). Additionally, there was no statistical difference between the BMI of the subjects in the MDT and the BMI of the other subjects in the sample (26.21 ± 3.90 kg/m^2 vs. 26.60 ± 4.52 kg/m^2; $t_{204} = -0.503$; $P = 0.616$).