Biodemography and Social Biology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsbi20

Development and Evaluation of a Genetic Risk Score for Obesity

Daniel W. Belsky [a b c], Terrie E. Moffitt [b d e], Karen Sugden [b d e], Benjamin Williams [b d e], Renate Houts [d], Jeanette McCarthy [f] & Avshalom Caspi [b d e]

[a] Department of Health Policy & Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[b] Institute for Genome Sciences & Policy, Duke University, Durham, North Carolina, USA

[c] Center for the Study of Aging and Human Development, Duke University Medical Center, Durham, North Carolina, USA

[d] Departments of Psychology and Neuroscience, Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina, USA

[e] Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, United Kingdom

[f] Department of Epidemiology & Biostatistics, University of California, San Francisco School of Medicine, San Francisco, California, USA

Published online: 23 May 2013.

To cite this article: Daniel W. Belsky, Terrie E. Moffitt, Karen Sugden, Benjamin Williams, Renate Houts, Jeanette McCarthy & Avshalom Caspi (2013): Development and Evaluation of a Genetic Risk Score for Obesity, Biodemography and Social Biology, 59:1, 85-100

To link to this article: http://dx.doi.org/10.1080/19485565.2013.774628

Biodemography and Social Biology, 59:85–100, 2013

ISSN: 1948-5565 print / 1948-5573 online
DOI: 10.1080/19485565.2013.774628

Routledge
Taylor & Francis Group

Development and Evaluation of a Genetic Risk Score for Obesity

DANIEL W. BELSKY,[1,2,3] TERRIE E. MOFFITT,[2,4,5] KAREN SUGDEN,[2,4,5] BENJAMIN WILLIAMS,[2,4,5] RENATE HOUTS,[4] JEANETTE McCARTHY,[6] AND AVSHALOM CASPI[2,4,5]

[1]Department of Health Policy & Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[2]Institute for Genome Sciences & Policy, Duke University, Durham, North Carolina, USA
[3]Center for the Study of Aging and Human Development, Duke University Medical Center, Durham, North Carolina, USA
[4]Departments of Psychology and Neuroscience, Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina, USA
[5]Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, United Kingdom
[6]Department of Epidemiology & Biostatistics, University of California, San Francisco School of Medicine, San Francisco, California, USA

Multi-locus profiles of genetic risk, so-called "genetic risk scores," can be used to translate discoveries from genome-wide association studies into tools for population health research. We developed a genetic risk score for obesity from results of 16 published genome-wide association studies of obesity phenotypes in European-descent samples. We then evaluated this genetic risk score using data from the Atherosclerosis Risk in Communities (ARIC) cohort GWAS sample (N = 10,745, 55% female, 77% white, 23% African American). Our 32-locus GRS was a statistically significant predictor of body mass index (BMI) and obesity among ARIC whites [for BMI, $r = 0.13$, $p<1 \times 10^{-30}$; for obesity, area under the receiver operating characteristic curve (AUC) = 0.57 (95% CI 0.55–0.58)]. The GRS predicted differences in obesity risk net of demographic, geographic, and socioeconomic information. The GRS performed less well among African

This research received support from US National Institute on Aging grant AG032282 and UK Medical Research Council grants G0100527 and G0601483. Additional support was provided by the Jacobs Foundation. Dr. Belsky was supported in part by US Agency for Healthcare Research and Quality grant 1R36HS020524-01 and by US National Institute on Aging grant T32-AG000029. The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C, R01HL087641, R01HL59367 and R01HL086694); National Human Genome Research Institute contract U01HG 004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. ARIC data (phs000090.v1.p1) were obtained from dbGaP.

Address correspondence to Daniel W. Belsky, 2020 W. Main St., Suite 201, Durham, NC 27708. E-mail: dbelsky@duke.edu

Americans. The genetic risk score we derived from GWAS provides a molecular measurement of genetic predisposition to elevated BMI and obesity.

[Supplemental materials are available for this article. Go to the publisher's online edition of Biodemography and Social Biology *for the following resource: Supplement to Development & Evaluation of a Genetic Risk Score for Obesity.]*

Introduction

Genome-wide association study (GWAS) results represent a potentially rich source of information for etiological and treatment research that builds bridges between genome science and clinical and public health practice (Janssens 2008; Khoury, McBride, et al. 2009). Given the large number of such studies, sufficient GWAS data exist to support such translational research for a number of common chronic health conditions, including obesity (Hindorff et al. 2009; Wray, Goddard, and Visscher 2007). Infrastructure is already in place at the start of the translational pipeline, with GWAS data banked and curated in continuously updated searchable databases (Hindorff et al. 2009; Yu et al. 2008). Likewise, at the other end of the pipeline, evidence from translational research is evaluated to establish the clinical utility of genomic information and issue guidelines for clinical practice (Khoury, Feero, et al. 2009). However, significant gaps remain in the middle of the translational pipeline, and approaches are needed to support research at this juncture, so that population-based samples with rich environmental and phenotypic measurements can be used to follow up disease markers identified in GWAS. Specifically, systematic approaches are needed to sift the results of numerous association studies and distill the most promising set of markers for further investigation. These approaches must be able to harness the power of existing resources and flexibly accommodate the rapid rate of data discovery in genome science.

A key hurdle for research using GWAS results is that risky single-nucleotide polymorphisms (SNPs identified in GWAS may not cause adverse health outcomes but may instead be proxies for (i.e., correlated with) unmeasured disease-causing variation in the genome (Gibson and Goldstein 2007; Orozco, Barrett, and Zeggini 2010). GWAS methods exploit linkage disequilibria (LD) across the genome to leverage the measurement of 100,000 to one million SNPs and capture variation in the 10 million–plus SNPs the genome is estimated to contain. The very large sample sizes in GWAS permit detection of risk associations even when proxy SNPs are in imperfect LD with disease-causing variation (correlation < 1). GWAS findings are generally applied to smaller samples designed to elucidate etiological and clinical correlates of discovered genes. When GWAS SNPs are translated to research using smaller samples, the measurement error resulting from imperfect LD with disease-causing variants can attenuate associations below levels these samples are powered to detect. Genetic risk scores (GRSs) summarize risk-associated variation across the genome (Horne et al. 2005) by aggregating information from multiple-risk SNPs (the simplest GRSs count disease-associated alleles). Because GRSs pool information from multiple SNPs, each individual SNP is less important to the summary measurement, and the "signal" from the GRS is robust to imperfect linkage for any one SNP. For the same reason, GRSs are less sensitive to minor allele frequencies for individual SNPs. As the number of SNPs included in a GRS grows, the distribution of values approaches normality, even when individual risk alleles are relatively uncommon (Fisher 1918). Therefore, the GRS can be an efficient and effective means of constructing genome-wide risk measurements from GWAS findings.

Obesity is a public health problem that is well suited to risk assessment using a GRS. It is highly prevalent (Ogden et al. 2006); it is a significant source of health care costs, morbidity, and mortality (Adams et al. 2006; Allender and Rayner 2007; Trogdon et al. 2008); it is under strong genetic influence (Yang, Kelly, and He 2007); and GWAS are beginning to elucidate its molecular genetic roots (O'Rahilly 2009). Therefore, translational research in obesity genomics may ultimately help to address a public health priority. A key challenge in this effort is that obesity's genetic roots are diffuse, multifactorial, and nondeterministic; many variants scattered across the genome each contribute small risks for obesity (McCarthy 2010). In other words, information from multiple genetic variants is needed to characterize genetic susceptibility to obesity. Thus, a GRS may be useful. A further challenge is uncertainty about the specific genetic variants that should be included in an obesity GRS. Different GWAS identify different genomic loci, and when loci are replicated across GWAS, the specific SNPs identified may be different (Hindorff et al. 2010). To address this challenge, we developed a three-stage approach to review GWAS results and select specific SNPs to include in a GRS. We devised our approach to be systematic and replicable and to leverage the discovery potential of GWAS while minimizing the risk of including false-positive markers. In this article, we describe this three-stage approach, apply it to develop a GRS for obesity, and test the GRS as a measure of obesity risk using data from the population-based Atherosclerosis Risk in the Communities (ARIC) Study.

Methods

Sample

The ARIC sample is described in detail elsewhere (Folsom et al. 2006; ARIC Investigators 1989). Briefly, ARIC is a prospective epidemiologic cohort study sponsored by the National Heart, Lung, and Blood Institute to investigate the etiology of atherosclerotic disease. The study draws from four U.S. communities: Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina; and Jackson, Mississippi. Participants were examined first during 1987–1989, and then during three subsequent periods (1990–1992, 1993–1995, and 1996–1998), with ongoing follow-up conducted annually by telephone. ARIC cohort genotype data from the Affymetrix Affy 6.0 Chip and selected phenotypes were obtained for this study from the NIH database of Genotypes and Phenotypes (dbGaP).

The original ARIC sample includes 15,792 participants (27% African American, 55% female). The publicly available dataset obtained from dbGaP for this study includes genotype and phenotype data for 12,771 individuals. Of this sample, 1,212 participants had a missing call rate that was greater than 2 percent for SNPs called successfully in greater than or equal to 95 percent of the sample and were excluded from subsequent analyses, following the quality control recommendations of the GENEVA ARIC Project (GENEVA ARIC Project 2009). In addition, although the ARIC study design did not aim to include relatives, genomic analysis by the ARIC investigators revealed familial relationships at the level of half-siblings or closer among 1,674 participants. One member was selected at random from each of the 105 "families" uncovered by this investigation to form a sample of unrelated persons. After these exclusions, the sample consisted of 10,745 participants (23% African American, 55% female, hereafter referred to as the "analysis sample").

Body Mass Index and Obesity

Body mass index (BMI) was calculated using measurements of weight rounded to the nearest pound and height rounded to the nearest centimeter. Obesity was defined as a BMI greater than or equal to 30, according to the criteria established by the U.S. Centers for Disease Control and Prevention. Anthropometric measurements were collected from participants wearing a scrub suit and no shoes at the four in-person data collection sessions.

Genotypes

Details on the genotyping of the ARIC sample are available through dbGaP and are described in detail elsewhere (Psaty et al. 2009). Briefly, genotyping was conducted by the Broad Institute using the Affymetrix Affy 6.0 SNP array and the Birdseed calling algorithm (Korn et al. 2008). Following guidelines for the use of genotypic data provided by the ARIC GWAS team, data were extracted for all SNPs with a sample-wide call rate greater than or equal to 95 percent, with fewer than five discordant calls across duplicated DNA samples in the quality-control subsample ($n = 334$) and in Hardy-Weinberg Equilibrium ($p > .001$).

Genetic Risk Scores

Current mid-pipeline translational studies use either a "best-guess" approach or a "top-hits" approach to select genetic markers to include in GRSs. The best-guess approach selects markers identified in association studies that are located in or near genes with plausible biological relationships to the pathophysiology of a phenotype or that demonstrate strong and replicable association signals (Lyssenko et al. 2008; Morrison et al. 2007; Talmud et al. 2010). The top-hits approach selects those markers with the strongest association signals in a single GWAS, independent of their biological plausibility (Demirkan et al. 2010; He et al. 2010). Early studies have illustrated the promise of translational research with GWAS markers, but as the field moves forward, more systematic approaches are needed that can better integrate new information from the latest studies. Neither the top-hits nor the best-guess approach provides a systematic and replicable means of integrating results from multiple GWAS. Meta-analysis can accomplish this integration, but comprehensive meta-analyses are not always available. Moreover, the top-hits and best-guess approaches do not provide a means to select specific SNPs for follow-up, and this problem is not solved by meta-analysis. The approach of selecting the "lead" SNP at a locus—usually the SNP with the lowest p value in the largest GWAS—is problematic, because different GWAS can report different lead SNPs for the same locus because of differences in GWAS chips, genotyping quality, and data-handling and analysis decisions. Thus, an approach is needed that facilitates systematic and replicable SNP selection from results of multiple GWAS.

Our three-stage approach integrates public-access resources including continuously updated databases of GWAS results, Web-based whole-genome analysis tools, and genome-wide data to identify the most promising set of SNPs for follow-up. Most important, the three-stage approach addresses key limitations of the top-hits and best-guess approaches, providing a systematic and replicable means of integrating findings across multiple GWAS and selecting SNPs for follow-up in new samples. The three stages are:

1. *Extraction*. All SNPs associated with one of the selected phenotypes at a given significance threshold are "extracted" from each GWAS and retained for further analysis.
2. *Clustering:*.Extracted SNPs are "clustered" according to patterns of LD that are determined from a reference population that matches the population in the GWAS included in Stage 1. Clustering yields a set of "LD blocks."
3. *Selection*. Statistical significance and replication are evaluated at the level of the LD block. The original GWAS results are used to assign a minimum *p* value and a replication count for each LD block. The minimum *p* value is the lowest *p* value reported for any SNP in the LD block in any GWAS contributing data in Stage 1. The replication count is the number of GWAS that reported an association for any SNP in the LD block at the threshold defined in Stage 1.

We applied our three-stage approach to construct two GRSs for obesity. First, we considered only GWAS published in print or online through December 31, 2008. We chose these GWAS because they were used in previous research that created top-hits and best-guess obesity GRSs. Thus, we used these GWAS to construct a GRS using our three-stage approach and then compared it to two previously published GRSs (Li et al. 2010; Peterson et al. 2011). Second, we considered all GWAS published through December 31, 2010. We applied our three-stage approach to the results from the full set of GWAS and compared the resulting GRS to a top-hits GRS generated from the largest meta-analysis of BMI GWAS published to date (Speliotes et al. 2010), as well as to a best-guess GRS generated from the full set of obesity-associated SNPs reported in the National Human Genome Research Institute (NHGRI) GWAS Catalog (Hindorff et al. 2010). The derivation of the GRS using the three-stage approach is described in detail in the supplemental material (see Supplemental Methods and Supplemental Tables 1–7). Analyses (also described in the supplemental material) revealed that the three-stage approach created GRSs that were at least as predictive of BMI and obesity as GRSs created with the top-hits and best-guess approaches. Further analyses to refine the GRS created through the three-stage approach yielded a final set of 32 SNPs. We applied two weighting schemes to the 32 SNPs before summing them to create our obesity GRS: (1) equal weighting, under which the score was a simple count of BMI-increasing alleles; and (2) effect-size weighting, under which BMI-increasing alleles were weighted by the effect size reported for that locus in the GIANT Consortium (Speliotes et al. 2010) or the DeCode BMI GWAS (Thorleifsson et al. 2009). Effect-size weights were adjusted for LD between the SNP tested in the GWAS and the SNP genotyped in the ARIC sample. Each of the 32 SNPs in the GRS was missing for fewer than 1 percent of participants in any gender/ethnicity cell. GRSs were prorated by dividing the GRS by the number of SNPs contributing data and then multiplying by 32. The SNPs included in the final obesity GRS, their BMI-increasing ("effect") alleles, nearby genes, and weights are reported in Table 1.

Evaluation of the Obesity GRS

Associations between the GRS and obesity-related traits (BMI, weight, waist circumference, obesity) were tested with linear and logistic regression models. These and subsequent models were adjusted for demographic and geographic control variables: age was specified as a linear and a quadratic term; a product term was included for the interaction between age and sex to account for sex differences in BMI and obesity distributions at different ages; and the four ARIC Study Centers where participants were enrolled in the study were

Table 1
Single nucleotide polymorphisms included in the obesity genetic risk score (GRS)

Chr	Nearby Gene	SNP	Effect Allele	Other Allele	Weight	Effect Allele Frequency (ARIC Sample)	
						Whites	African Americans
1	NEGR1	rs2815752	G	A	0.13	62%	55%
	TNNI3K	rs1514175	A	G	0.07	43%	68%
	PTBP2	rs1555543	A	C	0.06	58%	43%
	SEC16B	rs543874	G	A	0.22	20%	25%
2	FANCL	rs759250	A	G	0.10	29%	8%
	LRP1B	rs2121279	T	C	0.08	14%	3%
	TMEM18	rs2867123	G	C	0.30	83%	88%
	RBJ	rs10182181	G	A	0.14	54%	16%
3	CADM2	rs12714640	A	C	0.10	19%	6%
	ETV5/DGKG	rs1516728	T	A	0.11	77%	48%
4	GNPDA2	rs12641981	T	C	0.18	43%	23%
	SLC39A8	rs13114738	T	C	0.13	8%	1%
5	POC5 FLJ35779	rs10057967	C	T	0.10	63%	51%
	ZNF608	rs6864049	A	G	0.07	54%	81%
6	TFAP2B	rs734597	A	G	0.13	17%	9%
9	LING02 LRRN6C	rs1412235	C	G	0.11	31%	16%
	LMX1B	rs867559	G	A	0.24	20%	32%
11	RPL27A	rs2028882	C	A	0.06	50%	34%
	BDNF	rs10501087	C	T	0.18	79%	93%
	MTCH2	rs12419692	A	C	0.05	36%	9%
12	BDCDIN3D, FAIM2	rs7138803	A	G	0.12	38%	17%
13	MTIF3, GRF3A	rs1475219	C	T	0.09	21%	22%
14	PRKD1	rs1440983	A	G	0.15	5%	23%
	NRXN3	rs7144011	T	G	0.13	22%	24%
15	MAP2K5	rs28670272	G	A	0.13	77%	59%
16	GPR5B	rs11639988	G	A	0.17	85%	76%
	ATXN2L, TUFM, SH2B1	rs12443881	T	C	0.15	39%	9%
	FTO	rs9939609	A	T	0.38	41%	48%
18	MC4R	rs12970134	A	G	0.21	26%	13%
19	KCTD15	rs11084753	A	G	0.04	67%	64%
	QPCTL	rs11083779	C	T	0.07	96%	89%
	ZC3H4 TMEM160	rs7250850	G	C	0.09	71%	20%

Notes: Alleles are reported from the forward strand. The GRS was computed by counting the number of effect alleles at each SNP, multiplying that number by the SNP's weight, and then summing the results across the set of 32 SNPs. Weights reflect per-allele changes in BMI estimated in the the GIANT Consortium GWAS meta-analysis (Speliotes et al. 2010), except for rs867559, for which the weight was estimated in the DeCODE GWAS meta-analysis (Thorleifsson et al. 2009).

entered as a series of dummy variables (this collection of variables is referred to hereafter as "demographics and geography"). Predictiveness of the GRS was evaluated using three metrics that are established tools for evaluating risk markers in general (McGeechan et al. 2008), as well as for evaluating the specific case of genetic risk scores (Mihaescu et al. 2010). The first metric was R^2, the proportion of variation explained in BMI. The second metric used was AUC, the area under the receiver-operating characteristic (ROC) curve for obesity, also known as the discrimination index. R^2 was estimated using demographics and geography-adjusted linear regression models. The AUC corresponds to the probability that a randomly selected obese case will have a higher GRS than a randomly selected nonobese control. A marker that discriminates no better than chance has an AUC of 0.50. A marker that discriminates perfectly has an AUC of 1. A related metric is the partial AUC (PAUC), which sets a specificity threshold and calculates an AUC-like statistic for that specificity. Analyses of PAUC for the GRS set specificity at 80 percent (the bottom fifth of the ROC curve). AUC and PAUC analyses were stratified by the ARIC Study Center using Pepe's method (Janes and Pepe 2009). To determine whether the GRS improved discrimination over and above demographic and geographic information, we calculated a second set of statistics, delta AUC and delta PAUC. Probit regression models were used to generate predicted probabilities of obesity for each ARIC participant using a baseline model that included demographic and geographic information and a test model that also included the GRS. AUCs were calculated using these predicted probabilities as "risk scores" (Pepe, Cai, and Longton 2006), and estimates of the differences between the baseline and test models were bootstrapped to obtain confidence intervals. AUC analyses were conducted using the Stata package "comproc" (Pepe, Longton, and Janes 2009). 3) The third and final metric used was the integrated discrimination index (IDI) for obesity. The IDI evaluates the added predictiveness of a marker by comparing predictions made using a baseline set of risk markers to predictions that also include information about the new risk marker:

$$\text{IDI} = \left(\text{Prob}_{\text{test, obese}} - \text{Prob}_{\text{test,non}-\text{obese}}\right) - \left(\text{Prob}_{\text{baseline, obese}} - \text{Prob}_{\text{baseline, non}-\text{obese}}\right)$$

where "Prob" is the average predicted probability for a particular group from a particular model. The IDI measures change in model sensitivity net of change in model specificity and is a more sensitive measure than delta AUC (Pencina, D'Agostino, and Vasan 2008). An IDI of zero indicates that the test model performs comparably to the baseline model. Positive IDI values index net improvement in model sensitivity. Baseline and test models for IDI analyses were identical to those used in delta AUC analyses.

We tested differences between the predictiveness metrics for different risk scores by bootstrapping confidence intervals around the R^2 and AUC metrics (comparing the difference in estimated metric values across 1,000 random samples drawn with replacement from the ARIC database; see Pepe, Longton, and Janes 2009) and by applying Pencina's method (Pencina, D'Agostino, and Vasan 2008) to test change in the IDI metric. Comparisons were as follows: unweighted GRS versus weighted GRS; weighted GRS versus simple genetic risk assessment (the sum of risk alleles at the two best-replicated obesity loci: rs9939606, found in the gene *FTO*, and rs12970134, found downstream of the gene *MC4R*); weighted GRS versus socioeconomic index (educational attainment measured in six categories: grade school or less, some high school, high school graduate, vocational school, college, and graduate/professional school; information can be found in Supplementary Table 8).

Results

Obesity risk-allele distributions were similar for males and females but different for whites and African Americans. The variance of the unweighted GRS was greater for whites as compared to African Americans ($SD = 3.50$, as compared to 3.25, $p < .001$, using Brown and Forsythe's method; Brown and Forsythe 1974), as was the mean ($M = 28.80$, as compared to 24.87, $p < .001$, using a t-test for unequal variances; see also Supplementary Figure 1, available from the authors on request). This difference reflected lower frequencies of BMI-increasing alleles for several GRS SNPs among African American ARIC participants (see Table 1). Subsequent analyses were stratified by race.

The obesity GRSs were weakly but consistently associated with BMI and the probability of being obese among whites and African Americans, but associations were weaker among African Americans (see Figure 1). Among whites, after adjusting for age, sex, and geography, the unweighted GRS was associated with BMI at $r = 0.12$, and the weighted GRS was associated with BMI at $r = 0.13$ ($p < 1 \times 10^{-26}$ for both). This effect size corresponded to a 0.60-unit increase in BMI per standard deviation increase in the GRS. For each standard deviation increase in their unweighted and weighted GRSs, white ARIC participants' risk for obesity increased by 19.35 percent and 20.51 percent, respectively ($p < 1 \times 10^{-18}$ for both). Among African Americans, the weighted and unweighted GRSs were associated with BMI at $r = 0.05$ ($p < .05$ for both). For each standard deviation increase in their unweighted and weighted GRSs, African American ARIC participants'

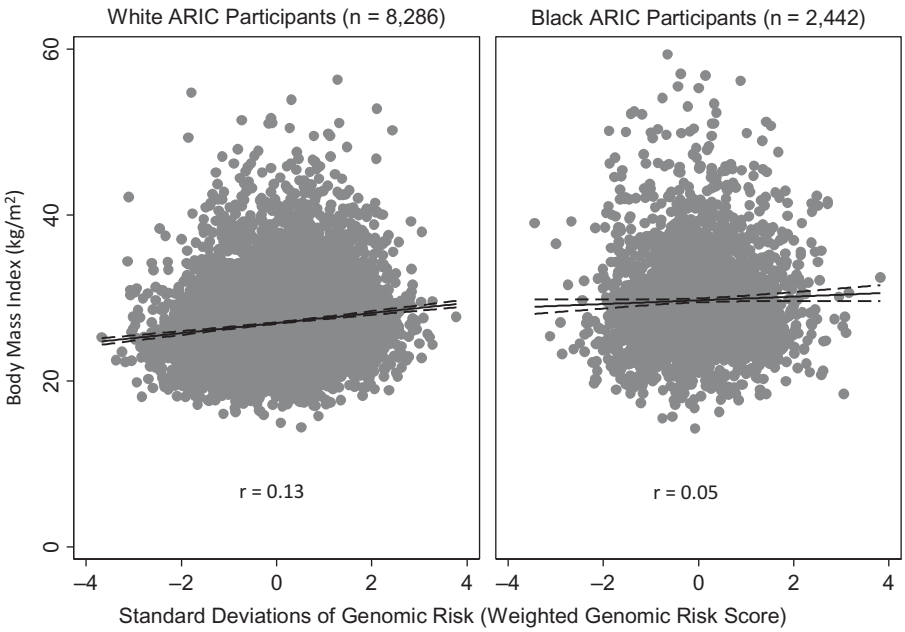


Figure 1. Panel A. BMI for African American and white ARIC participants plotted against the weighted obesity genetic risk score. Dashed outlines represent 95 percent confidence intervals. Pearson correlations (r) were adjusted for gender, age, and data collection at the ARIC Study Center. Removal of outliers (not shown) did not alter correlation estimates at the third decimal point. Correlations were statistically significant for white ($p < 1 \times 10^{-30}$) and African American ($p = .014$) ARIC participants.

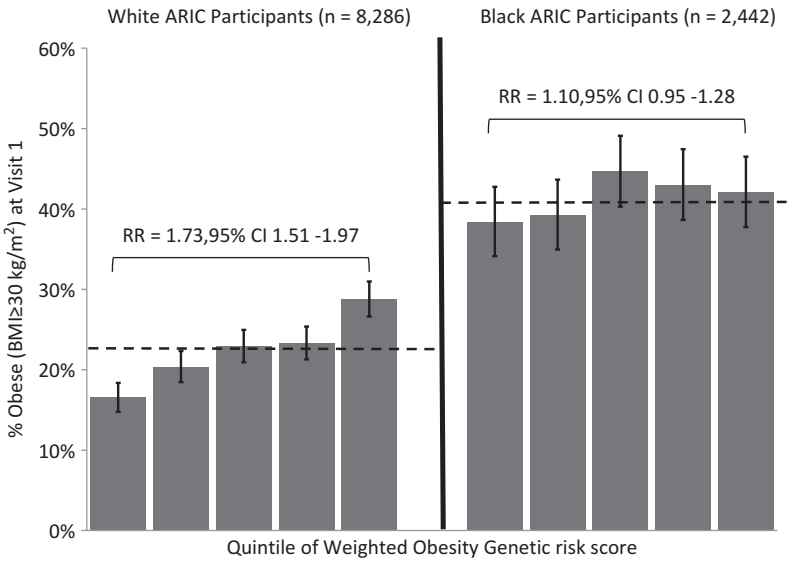


Figure 1. Panel B. Percentage white and African American ARIC participants who were obese (BMI ≥ 30) at the first study visit, by quintile of genetic risk score. Quintiles were determined separately for whites and African Americans. Error bars represent 95 percent confidence intervals. Risk ratios are for comparisons of highest to lowest quintiles of genomic risk and were estimated with adjustment for gender, age, and data collection at the ARIC Study Center. Dashed lines represent sample means. Among white ARIC participants, all quintile-to-quintile differences are statistically significant ($p <$ .01), with the exception of the third and fourth quintiles. Among African American ARIC participants, the percentage of obese individuals in the lowest quintile was lower than the percentage of obese individuals in the third and fourth quintiles ($p < .05$).

risk for obesity increased by 3.54 percent ($p = .059$) and 4.92 percent ($p = .017$), respectively. Results were substantively unchanged when control variables were removed from the models.

Although we stratified our analyses by ethnicity (whites and blacks), we conducted an additional analysis to determine whether population stratification within the white and black subsamples influenced our estimates of GRS-BMI or GRS-obesity associations. Principal components derived from eigen analysis of genome-wide SNP data can be used to control for population stratification in genetic association analyses (Price et al. 2006). Such principal components were derived separately for whites and blacks in the ARIC cohort using the method described by Patterson, Price, and Reich (2006) and were included in the database we obtained from dbGaP (GENEVA ARIC Project 2009). Adjustment for these principal components as covariates in regression analyses did not change our results.

We conducted a series of additional sensitivity analyses to evaluate heterogeneity in GRS associations (described in detail in the supplementary material, available from the authors on request). These analyses supported a linear association between the GRS and BMI, showed that GRS-BMI associations were similar to GRS-weight and GRS–waist circumference associations, and revealed no sex or age differences in GRS-BMI associations.

The obesity GRSs performed similarly on the three predictiveness metrics (see Table 2). The top panel of Table 2 addresses clinical validity and presents the three metrics

Table 2
Predictiveness metrics for the three-stage approach obesity genetic risk score and comparison measures of risk for obesity

	White ARIC Participants (n = 8,286)			Black ARIC Participants (n = 2,442)		
	R^2 (95% CI)	AUC (95% CI)	IDI (p-value)	R^2 (95% CI)	AUC (95% CI)	IDI (p-value)
Panel A. Predictivness of the un-weighted and weighted obesity GRSs						
Un-Weighted GRS	1.39%	0.565	0.009	%0.11	0.515	0.001
	(0.94%−1.89%)	(0.550−0.581)	(4.65E−18)	(−%0.04−%0.57)	(0.491−0.540)	(0.067)
Weighted GRS	1.57%	0.570	0.010	%0.14	0.521	0.002
	(1.11%−2.10%)	(0.554−0.584)	(8.25E−20)	(−%0.03−%0.65)	(0.497−0.544)	(0.152)
Panel B. Predictiveness of comparison risk measures						
Simple Genetic Risk Assessment:	0.59%	0.543	0.004	−%0.02	0.516	0.001
FTO & MC4R-linked SNPs only	(0.31%−0.97%)	(0.528−0.557)	(3.54E−09)	(−%0.04−%0.25)	(0.493−0.539)	(0.149)
Socioeconomic Index: 6-category	0.57%	0.532	0.003	%1.06	0.561	0.016
measure of educational attainment	(0.29%−0.87%)	(0.517−0.546)	(7.83E−07)	(%0.42−%1.99)	(0.538−0.584)	(2.71E−11)
Panel C. Predictiveness of model-based risk assessments (including demographic and geographic information)						
Simple Genetic Risk Assessment	3.88%	0.550		5.35%	0.607	
Weighted GRS (includes simple genetic risk assessment)	4.88%	0.574		5.52%	0.609	
Change in predictiveness with	1.00%	0.024	0.006	0.17%	0.002	0.001
addition of weighted GRS to model	(0.58% − 1.42%)	(0.012 − 0.036)	(7.81E−13)	(−%0.15 − %0.51)	(−0.005 − 0.009)	(0.055)
Socioeconomic Status	4.70%	0.550		7.70%	0.643	
Socioeconomic Status + weighted GRS	6.20%	0.586		7.92%	0.645	
Change in predictiveness with	1.50%	0.036	0.010	0.22%	0.002	0.002
addition of weighted GRS to model	(1.00% − 1.99%)	(0.023 − 0.050)	(5.46E−19)	(−%0.14 − %0.55)	(−0.003 − 0.008)	(0.012)

Notes: The simple genetic risk score is a component of the weighted obesity genetic risk scores. Values of R^2 were estimated using linear regression models adjusted for demographics and geography. Percentile-based confidence intervals were generated using the bootstrap method. AUCs and percentile-based confidence intervals were estimated from ROC curves constructed for raw values (i.e., actual values of the measures tested, rather than predicted values generated from a regression model) and were adjusted for the geography where data were collected. IDIs and test statistics were estimated using comparisons of a baseline model that included demographic and geographic information to a test model that included both this information and the GRS.

for the unweighted and weighted GRSs. Among whites, weighted and unweighted obesity GRSs explained small but statistically significant proportions of the variance in BMI (R^2), discriminated obese from nonobese participants modestly better than chance (AUC), and contributed small net improvements to the sensitivity of an obesity prediction model over and above demographic and geographic information (IDI). Among African Americans, the GRS did not contribute to the explanation of variance in BMI over and above demographic and geographic information, to the discrimination of obese from nonobese participants, or to the net sensitivity of the obesity prediction model. Use of weights derived from BMI GWAS improved the performance of the GRS among whites and African Americans, but this improvement was not statistically significant ($p > .10$ for all comparisons).

The bottom panel of Table 2 addresses research utility, presenting predictiveness metrics for two comparison measures of obesity risk: the simple genetic risk assessment (weighted combinations of rs9939609 in *FTO* and rs12970134 downstream of *MC4R*) and the socioeconomic index (a six-category measure of educational attainment). The *FTO* and *MC4R* loci and socioeconomic status are robust correlates of BMI and obesity in adult samples (Ford and Mokdad 2008; Hardy et al. 2010). Comparison of the 32-locus GRS to a two-locus risk assessment can illustrate whether the GRS offers value added over a simpler genetic risk assessment. Comparison of the GRS to socioeconomic status can illustrate how the predictiveness of the GRS compares to the predictiveness of a social determinant of obesity that is not easily changed but is understood to be important in etiological research (Drewnowski 2009). Among whites, the genetic risk scores performed better than the comparison measures of obesity risk on all three metrics ($p < .01$ for all comparisons). Among African Americans, the GRSs performed no differently than the simple genetic risk assessment ($p > .10$) and performed less well than the socioeconomic index ($p = .021$). When combined with the comparison risk measures and demographic and geographic information, the GRS improved predictiveness for whites but not for African Americans (see Supplementary Table 9, available from the authors on request).

Figure 2 shows the model-based ROC curves for a baseline model that included demographic and geographic information and a test model that also included the weighted GRS. The change in AUC from the baseline model to the test model was greater than zero (delta AUC = 0.048, 95% CI = 0.313–0.658, $p < 10^{-7}$), indicating that the GRS improved discrimination of obese cases. This improvement in discrimination was concentrated at low specificities but did extend to the portion of the ROC curve that is of greatest interest to clinicians. At a specificity of 0.8, the test model including the GRS was marginally more sensitive than the baseline model (delta PAUC = 0.007, 95% CI < 0.0003–0.010, $p < .001$). Results for African Americans are presented in Supplementary Figure 2 (available from the authors on request).

Discussion

We used a three-stage approach to construct an obesity GRS from GWAS results. Our tests of this obesity GRS in the population-based ARIC cohort revealed it to be a highly statistically significant predictor of BMI, as measured at four time points across 10 years; weight and waist circumference; and obesity. In terms of value added, the GRS improved prediction of BMI and obesity over and above demographic and geographic information, *FTO* and *MC4R* genotypes, and information about socioeconomic status. Thus, the GRS provides a measure of genetic predisposition to obesity that could inform etiological and treatment research.

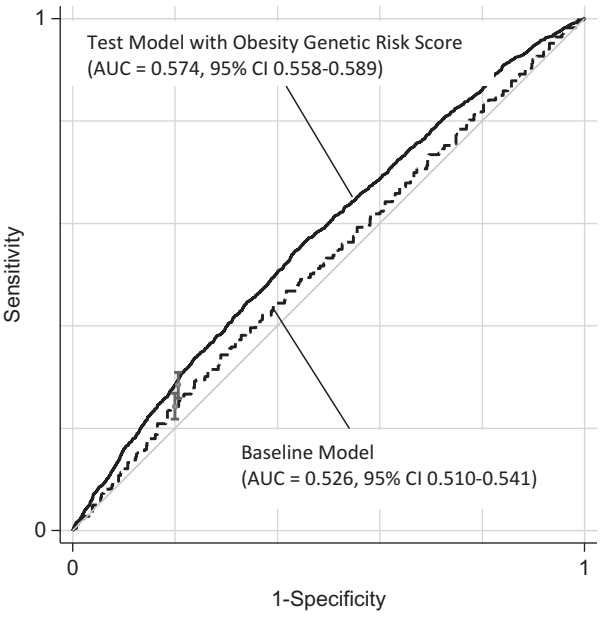


Figure 2. Receiver-operating characteristic (ROC) curves for obesity among white ARIC participants (n = 8,286). Baseline model = gender, age (quadratic), gender-age interaction, ARIC Study Center; Test Model = baseline model + weighted obesity genetic risk score. ROC curves were constructed using predicted values from probit regressions of obesity (BMI ≥ 30) on the model terms. Delta AUC ($AUC_{Test} - AUC_{Baseline}$) = 0.048, 95% CI = 0.031–0.066, $p < 1 \times 10^{-7}$. Delta PAUC at 80% specificity = 0.007, 95% CI = 0.003–0.010, $p < .001$. AUCs, PAUCs, and delta AUCs were estimated using Pepe's method (Janes and Pepe 2009; Pepe, Longton, and Janes 2009).

The research utility of the GRS is likely limited to samples of European descent. GRS-BMI and GRS-obesity associations for African American ARIC participants were much smaller than comparable associations for white ARIC participants. Although the sample included fewer African Americans than whites, power to detect effects of equal size as those observed in whites was well over 80 percent in the African American sample. Moreover, effect-size measures (r, R^2, relative risk, AUC, IDI) showed little evidence that the GRS predicted BMI or obesity among African Americans. These results suggest caution in using GWAS of European-descent populations to derive GRSs for African Americans. Our analyses did indicate that the GRS performed similarly among men and women. However, emerging evidence for gene-sex interactions in obesity (Benjamin et al. 2011; Heid et al. 2010) suggests that future obesity GRSs may require sex-specific construction.

Our results have implications for theory, research, and clinical practice. With respect to theory, our results are consistent with the hypothesis that genetic risk for obesity is quantitatively distributed and can be operationalized in a GRS (Plomin, Haworth, and Davis 2009). With respect to research methods, our findings illustrate one approach to operationalize quantitative genetic risk. A systematic and replicable approach to selecting SNPs from association studies to follow-up on etiological and treatment research will be especially important with the advent of next-generation sequencing approaches. Next-generation sequencing is likely to uncover many new disease-associated loci for obesity, as well as for other phenotypes of interest to clinicians and researchers. These variants, though rarer in the population, may have higher penetration and thus greater clinical relevance.

Future research can also make use of the GRS derived in this study as a measure of inherited obesity risk. With respect to clinical practice, results indicate that for persons in middle age, GWAS SNP-based approaches to obesity risk assessment offer little in the absence of more detailed information about lifestyle and environment. Although genetic information reliably predicted risk for obesity over and above demographics and geography, the magnitude of this additional risk was insufficient to recommend our score for use in clinical risk assessments. This finding is especially important in the context of questions regarding consumer genomics services (Evans et al. 2011). Our three-stage approach produced a more comprehensive genetic risk assessment for obesity than those currently produced by companies marketing genomics services directly to consumers. The very modest risk information furnished by our GRS recommends that health professionals be cautious in interpreting risk information provided by consumer genomics companies. The standard of evidence used here—multimethod assessment of predictiveness in large, population-based samples—should be considered a minimum standard for determining the validity of such risk information.

Results should be considered in light of the following limitations. First, some ARIC participants were included in the samples of some of the GWAS used to construct the GRS. However, these ARIC participants represented a minority of the GWAS samples, and results in the ARIC sample are similar to results from samples not included in any of the GWAS (Li et al. 2010; Peterson et al. 2011). Second, some risk loci identified by our three-stage approach could only be genotyped in the ARIC sample using relatively weak proxies. Given the small improvement to predictiveness associated with each additional SNP included in the GRS, it is unlikely that this limitation influenced the substance of our results, but it is possible that our GRS is moderately more predictive than analyses in the ARIC cohort suggest. Third, our analyses were limited to African Americans and white Americans. The ARIC cohort does not contain Asian or Hispanic individuals. It remains unclear whether the relatively greater similarity between these and European populations (Jorde and Wooding 2004) would support the generalization of our GRS. However, GWAS of Asian and Hispanic samples (He et al. 2010; Norris et al. 2009) suggest that a European-descent population-derived GRS may omit important risk loci for these populations. As more GWAS of non-European populations become available, our three-stage approach can be used to derive additional population-specific GRSs. Fourth, there is mounting evidence that many genetic factors predisposing individuals to obesity are sex-specific (McCarthy et al. 2003), and that GWAS that fail to model such sex-specificity may not detect important risk variants (McCarthy 2007). Results from GWAS modeling gene-sex interaction support this hypothesis (Benjamin et al. 2011; Chiu et al. 2010; Heid et al. 2010). As more such GWAS become available, our three-stage approach can be used to derive sex-specific GRSs for obesity. Finally, the ARIC sample is limited to individuals in middle age. There is evidence that genetic risk for obesity has dynamic consequences throughout development (Elks et al. 2010; Sovio et al. 2011). It will be important in subsequent investigations to evaluate our obesity GRS in longitudinal cohorts that capture a broader section of the life course, and particularly in young people, as they are a key prevention target (Belsky et al. 2012; Dietz 2004).

In this study, we constructed a GRS for obesity and showed that it predicted BMI and obesity in a population-based sample of middle-aged adults. These associations suggest that future research into obesity etiology and treatment can make use of genetic information. However, our analyses do not support the use of genetic testing for individual-level obesity-risk prediction. Future research with this GRS should characterize the expression of genetic

risk across the life course and particularly during childhood, when intervention to prevent the development of obesity may be most effective.

References

Adams, K. F., A. Schatzkin, T. B. Harris, V. Kipnis, T. Mouw, R. Ballard-Barbash, A. Hollenbeck, and M. F. Leitzmann. 2006. Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N Engl J Med* 355(8): 763–778.

Allender, S., and M. Rayner. 2007. The burden of overweight and obesity-related ill health in the UK. *Obes Rev* 8(5): 467–473.

ARIC Investigators. 1989. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol* 129(4): 687–702.

Belsky, D. W., T. E. Moffitt, R. Houts, G. G. Bennett, A. K. Biddle, J. A. Blumenthal, J. P. Evans, et al. 2012. Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a four-decade longitudinal study. *Arch Pediatr Adolesc Med* 166(6): 515–521.

Benjamin, A. M., S. Suchindran, K. Pearce, J. Rowell, L. F. Lien, J. R. Guyton, and J. J. McCarthy. 2011. Gene by sex interaction for measures of obesity in the framingham heart study. *J Obes* 2011:329, 038.

Brown, M. B., and A. B. Forsythe. 1974. Robust tests for equality of variances. *J Am Stat Assoc* 69: 364–367.

Chiu, Y. F., L. M. Chuang, H. Y. Kao, K. C. Shih, M. W. Lin, W. J. Lee, T. Quertermous, et al. 2010. Sex-specific genetic architecture of human fatness in Chinese: the SAPPHIRe Study. *Hum Genet* 128(5): 501–513.

Demirkan, A., B. W. Penninx, K. Hek, N. R. Wray, N. Amin, Y. S. Aulchenko, R. van Dyck, et al. 2011. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry* 16(7): 773–783.

Dietz, W. H. 2004. Overweight in childhood and adolescence. *N Engl J Med* 350(9): 855–857.

Drewnowski, A. 2009. Obesity, diets, and social inequalities. *Nutr Rev* 67(suppl 1): S36–S39.

Elks, C. E., R. J. Loos, S. J. Sharp, C. Langenberg, S. M. Ring, N. J. Timpson, A. R. Ness, et al. 2010. Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med* 7(5): e1000284.

Evans, J. P., E. M. Meslin, T. M. Marteau, and T. Caulfield. 2011. Genomics. Deflating the genomic bubble. *Science* 331(6019): 861–862.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans Royal Soc Edinburgh* 52: 399–433.

Folsom, A. R., L. E. Chambless, C. M. Ballantyne, J. Coresh, G. Heiss, K. K. Wu, E. Boerwinkle, et al. 2006. An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the Atherosclerosis Risk in Communities Study. *Arch Intern Med* 166(13): 1368–1373.

Ford, E. S., and A. H. Mokdad. 2008. Epidemiology of obesity in the Western hemisphere. *J Clin Endocrinol Metabol* 93(11 suppl 1): S1–S8.

GENEVA ARIC Project. 2009. Quality control report for the ARIC GWAS database. Bethesda, MD: National Institutes of Health Database of Genotypes and Phenotypes (dbGaP).

Gibson, G., and D. B. Goldstein. 2007. Human genetics: the hidden text of genome-wide associations. *Curr Biol* 17(21): R929–R932.

Hardy, R., A. K. Wills, A. Wong, C. E. Elks, N. J. Wareham, R. J. F. Loos, D. Kuh, and K. K. Ong. 2010. Life course variations in the associations between FTO and MC4R gene variants and body size. *Hum Mol Genet* 19(3): 545–552.

He, M. A., M. C. Cornelis, P. W. Franks, C. L. Zhang, F. B. Hu, and L. Qi. 2010. Obesity genotype score and cardiovascular risk in women with type 2 diabetes mellitus. *Arterioscl Thromb Vasc Biol* 30(2): 327–370.

Heid, I. M., A. U. Jackson, J. C. Randall, T. W. Winkler, L. Qi, V. Steinthorsdottir, G. Thorleifsson, et al. 2010. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 42(11): 949–960.

Hindorff, L. A., J. MacArthur, A. Wise, H. A. Junkins, P. N. Hall, A. K. Klemm, and T. A. Manolio. 2010. A catalog of published genome-wide association studies. Bethesda, MD: National Human Genome Research Institute. www.genome.gov/gwastudies.

Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106(23): 9362–9367.

Horne, B. D., J. L. Anderson, J. F. Carlquist, J. B. Muhlestein, D. G. Renlund, T. L. Bair, R. R. Pearson, and N. J. Camp. 2005. Generating genetic risk scores from intermediate phenotypes for use in association studies of clinically significant endpoints. *Ann Hum Genet* 69: 176–186.

Janes, H., and M. S. Pepe. 2009. Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika* 96(2): 371–382.

Janssens, A. C. J. W. 2008. Is the time right for translation research in genomics? *Eur J Epidemiol* 23(11): 707–710.

Jorde, L. B., and S. P. Wooding. 2004. Genetic variation, classification and "race." *Nat Genet* 36(11 suppl): S28–S33.

Khoury, M. J., W. G. Feero, M. Reyes, T. Citrin, A. Freedman, D. Leonard, W. Burke, et al. 2009. The genomic applications in practice and prevention network. *Genet Med* 11(7): 488–494.

Khoury, M. J., C. M. McBride, S. D. Schully, J. P. A. Ioannidis, W. G. Feero, A. C. J. W. Janssens, M. Gwinn, et al. 2009. The scientific foundation for personal genomics: recommendations from a National Institutes of Health–Centers for Disease Control and Prevention multidisciplinary workshop. *Genet Med* 11(8): 559–567.

Korn, J. M., F. G. Kuruvilla, S. A. McCarroll, A. Wysoker, J. Nemesh, S. Cawley, E. Hubbell, et al. 2008. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 40(10): 1253–1260.

Li, S., J. H. Zhao, J. Luan, R. N. Luben, S. A. Rodwell, K. T. Khaw, K. K. Ong, N. J. Wareham, and R. J. Loos. 2010. Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *Am J Clin Nutr* 91(1): 184–190.

Lyssenko, V., A. Jonsson, P. Almgren, N. Pulizzi, B. Isomaa, T. Tuomi, G. Berglund, D. Altshuler, P. Nilsson, and L. Groop. 2008. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359(21): 2220–2232.

McCarthy, J. J. 2007. Gene by sex interaction in the etiology of coronary heart disease and the preceding metabolic syndrome. *Nutr Metab Cardiovasc Dis* 17(2): 153–161.

——, 2010. Genomics, type 2 diabetes, and obesity. *N Engl J Med* 363(24): 2339–2350.

McCarthy, J. J., J. Meyer, D. J. Moliterno, L. K. Newby, W. J. Rogers, and E. J. Topol. 2003. Evidence for substantial effect modification by gender in a large-scale genetic association study of the metabolic syndrome among coronary heart disease patients. *Hum Genet* 114(1): 87–98.

McGeechan, K., P. Macaskill, L. Irwig, G. Liew, and T. Y. Wong. 2008. Assessing new biomarkers and predictive models for use in clinical practice: a clinician's guide. *Arch Intern Med* 168(21): 2304–2310.

Mihaescu, R., M. van Zitteren, M. van Hoek, E. J. Sijbrands, A. G. Uitterlinden, J. C. Witteman, A. Hofman, M. G. Hunink, C. M. van Duijn, and A. C. Janssens. 2010. Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. *Am J Epidemiol* 172(3): 353–361.

Morrison, A. C., L. A. Bare, L. E. Chambless, S. G. Ellis, M. Malloy, J. P. Kane, J. S. Pankow, J. J. Devlin, J. T. Willerson, and E. Boerwinkle. 2007. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* 166(1): 28–35.

Norris, J. M., C. D. Langefeld, M. E. Talbert, M. R. Wing, T. Haritunians, T. E. Fingerlin, A. J. Hanley, et al. 2009. Genome-wide association study and follow-up analysis of adiposity traits in Hispanic Americans: the IRAS Family Study. *Obesity* 17(10): 1932–1941.

Ogden, C. L., M. D. Carroll, L. R. Curtin, M. A. McDowell, C. J. Tabak, and K. M. Flegal. 2006. Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA* 295(13): 1549–1555.

O'Rahilly, S. 2009. Human genetics illuminates the paths to metabolic disease. *Nature* 462(7271): 307–314.

Orozco, G., J. C. Barrett, and E. Zeggini. 2010. Synthetic associations in the context of genome-wide association scan signals. *Hum Mol Genet* 19(R2): R137–R144.

Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. *PLoS Genet* 2(12): e190.

Pencina, M. J., R. B. D'Agostino, and R. S. Vasan. 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 27(2): 157–172.

Pepe, M. S., T. X. Cai, and G. Longton. 2006. Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics* 62(1): 221–229.

Pepe, M., G. Longton, and H. Janes. 2009. Estimation and comparison of receiver operating characteristic curves. *Stata J* 9(1):1.

Peterson, R. E., H. H. Maes, P. Holmans, A. R. Sanders, D. F. Levinson, J. Shi, K. S. Kendler, P. V. Gejman, and B. T. Webb. 2011. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Hum Genet* 129(2): 221–230.

Plomin, R., C. M. A. Haworth, and O. S. P. Davis. 2009. Common disorders are quantitative traits. *Nat Rev Genet* 10(12): 872–878.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8): 904–909.

Psaty, B. M., C. J. O'Donnell, V. Gudnason, K. L. Lunetta, A. R. Folsom, J. I. Rotter, A. G. Uitterlinden, T. B. Harris, J. C. Witteman, and E. Boerwinkle. 2009. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ Cardiovasc Genet* 2(1): 73–80.

Sovio, U., D. O. Mook-Kanamori, N. M. Warrington, R. W. Lawrence, L. Briollais, C. N. A. Palmer, J. Cecil, et al. 2011. Association between common variation at the FTO locus and changes in body mass index from infancy to late childhood: the complex nature of genetic association through growth and development. *PLoS Genet* 7(2): e1001307.

Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, et al. 2010. Association analyses of 249, 796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42(11): 937–948.

Talmud, P. J., A. D. Hingorani, J. A. Cooper, M. G. Marmot, E. J. Brunner, M. Kumari, M. Kivimaki, and S. E. Humphries. 2010. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* 340: b4838.

Thorleifsson, G., G. B. Walters, D. F. Gudbjartsson, V. Steinthorsdottir, P. Sulem, A. Helgadottir, U. Styrkarsdottir, et al. 2009. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41(1): 18–24.

Trogdon, J. G., E. A. Finkelstein, T. Hylands, P. S. Dellea, and S. J. Kamal-Bahl. 2008. Indirect costs of obesity: a review of the current literature. *Obes Rev* 9(5): 489–500.

Wray, N. R., M. E. Goddard, and P. M. Visscher. 2007. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 17(10): 1520–1528.

Yang, W. J., T. Kelly, and J. He. 2007. Genetic epidemiology of obesity. *Epidemiol Rev* 29: 49–61.

Yu, W., M. Gwinn, M. Clyne, A. Yesupriya, and M. J. Khoury. 2008. A navigator for human genome epidemiology. *Nat Genet* 40(2): 124–125.

Supplement

This supplement describes the application of the 3-stage approach to create a genetic risk score (GRS) for obesity. The supplement is organized into 3 sections: The first section describes the creation of the obesity GRS: Stage 1. Extraction; Stage 2. Clustering; and Stage 3. Selection. The second section describes analyses comparing the resulting GRS to GRSs created with the best-guess and top-hits approaches. The final section describes sensitivity analyses to test heterogeneity in GRS associations.

PART 1. CREATING THE OBESITY GRS

Stage 1. Extraction

For our 3-stage approach analyses, we considered GWAS of European-descent samples that targeted 4 phenotypes: obesity, weight, waist circumference, and body mass index (BMI) (hereafter "obesity-related phenotypes"). A search of the NGHRI GWAS Catalog using the HuGE Navigator (http://www.hugenavigator.org) identified 16 GWAS that met these inclusion criteria, 9 of which were published by December 31, 2008 (**Supplementary Table 1**).

In Stage 1 (Extraction), we compiled association results reported in the manuscripts and supplementary materials of the GWAS and extracted rs-numbers and p-values for SNPs associated with any of the 4 phenotypes in the discovery or combined discovery and replication samples at an alpha level of $1x10^{-5}$ (n=103 SNPs in the subset of 9 GWAS, n=519 SNPs in the full set of 16 GWAS, **Supplementary Table 2**). The significance level of $p<1x10^{-5}$ was the most generous threshold at which most GWAS published results and is the threshold used in the NHGRI GWAS Catalog (Hindorff et al. 2009). Associations were not extracted from replication samples because few GWAS reported novel associations identified in replication samples and some GWAS did not include replication samples or included replication samples of different ethnicity. Discovery sample risk SNPs that failed to replicate within an individual GWAS were included because replication was evaluated at the level of the GWAS publication rather than the specific test sample.

Stage 2. Clustering

In Stage 2 (Clustering), we grouped the extracted SNPs into "LD blocks." We defined LD blocks using data from the HapMap CEU sample (Phase 3), queried using Seattle SNPs' web-based Genome Variation Server (http://gvs.gs.washington.edu/GVS). For each SNP extracted in Stage 1 ("seeds"), we defined an LD block as the region containing all SNPs in LD with that seed at a threshold of $R^2 \geq 0.95$. Then, beginning

with the block closest to the start of each chromosome, we pruned blocks that did not contain a unique seed. This process yielded n=66 LD blocks from the subset of 9 GWAS published by December 31, 2008 and n=158 LD blocks from the full set of 16 GWAS.

Stage 3. Selection

In Stage 3 (Selection), we retained LD blocks that we classified as genome-wide significant or as replicated. Genome-wide significant LD blocks were those that contained ≥1 SNP associated with an obesity-related phenotype at $p<1x10^{-8}$. Replicated blocks were those that contained SNPs extracted from ≥2 GWAS. This process yielded n=37 LD blocks clustered around 11 loci on chromosomes 1-4,9,11,12,16,18, and 19 from the subset of 9 GWAS and n=69 LD blocks clustered around 32 loci on chromosomes 1-6,9,11-14,16,18, and 19 from the full set of 16 GWAS (**Supplementary Tables 3, 4**). Sensitivity analyses relaxing the LD threshold used to define LD blocks yielded fewer LD blocks (e.g., for the full set of 16 GWAS, n=58 at an R^2 threshold of 0.70), but did not alter the loci identified as genome-wide significant or replicated in the original analyses.

PART 2. COMPARING THE 3-STAGE APPROACH GRSS TO THE TOP-HITS AND BEST-GUESS GRSS

To construct and test our GRSs, we followed-up the LD blocks identified in our 3-stage approach analyses in the GWAS dataset from the Atherosclerosis Risk in Communities (ARIC) Study. This dataset is publicly available through the National Institutes of Health Database of Genotypes and Phenotypes (dbGaP) (http://www.ncbi.nlm.nih.gov/gap, phs000090.v1.p1) and is described in the Data section of the main text.

We selected SNPs in the ARIC database to include in our two GRSs as follows: We defined tag SNPs for each of the LD blocks as SNPs that were in LD with every seed contained in the block at $R^2 \geq 0.95$. We then matched 1 tag SNP per LD block with a SNP in the ARIC study genotype database that met the GENEVA ARIC Project Team's quality control criteria (GENEVA ARIC Project 2009). If no tag SNPs in an LD block could be matched in the ARIC database, we relaxed the LD threshold used to define a tag SNP until either a) the resulting set of tag SNPs overlapped with tag SNPs that we had already matched in the ARIC database, or b) a match with a new SNP in the ARIC database was achieved. These analyses yielded a set of n=28 SNPs from the subset of 9 GWAS and a set of n=57 SNPs from the full set of 16 GWAS.

To compute the 3-stage approach GRSs for each ARIC participant, we (1) identified the obesity-associated allele for each SNP from the GWAS where that SNP was reported; (2) calculated the mean number of risk alleles at each locus; and (3) summed these means across loci to produce the 3-stage approach genome-wide scores.

To compute the top-hits and best-guess approach GRSs, we selected SNPs from the ARIC database to match SNPs from 3 published GRSs (Li et al. 2010; Peterson et al. 2011; Speliotes et al. 2010) and the full set of obesity-associated SNPs listed in the NHGRI GWAS catalog for GWAS of European-descent samples. In cases where a specific SNP was not available in the ARIC database, we selected its closest LD proxy. We then summed obesity-associated alleles across each set of selected SNPs to create the comparison genome-wide scores.

To test if the 3-stage approach could construct a GRS that was at least as predictive of BMI and obesity as GRSs created with the top-hits and best-guess approaches, we compared effect sizes for different GRSs using the ARIC data. All GRSs were standardized to have mean=0 and standard deviation=1. To measure GRS effect sizes for BMI, we estimated Pearson correlations (r) from separate linear regressions of BMI on each of the GRSs. To measure GRS effect sizes for obesity, we estimated odds ratios (OR) from separate logistic regressions of obesity on each of the GRSs. Regression models were adjusted for age (linear and quadratic terms), gender, the age-gender interaction, and the ARIC Study Centers where data were collected (hereafter these statistical adjustments are described as "demographics and geography"). To test differences between GRS effect sizes, we conducted F-tests (for effect sizes estimated from linear regressions) and Wald tests (for effect sizes estimated from logistic regressions). For these tests, models including each of the GRSs being compared were jointly estimated using the seemingly unrelated regression method. Seemingly unrelated regression is a statistical approach for comparing coefficients from non-nested regression models (Baltagi 1980; Verzilli, Stallard, and Whittaker 2005). Effect sizes were similar for all GRSs. Statistical tests indicated that our 3-stage approach GRSs performed as well as or better than GRSs created using top-hits and best-guess approaches (**Supplementary Table 5**). Thus, the 3-stage approach produced a GRS that was at least as predictive as top-hits and best guess approach GRSs. We used the 3-stage approach GRS created from the full set of 16 GWAS (hereafter the "Obesity GRS") in subsequent analyses.

Refining the 3-Stage Approach GRS for Obesity. At 7 of the 32 loci identified in the 3-stage approach analyses of GWAS results (in or near the genes *TMEM18*, *ETV5*, *BDNF*, *MTCH2*, *FTO*, *MC4R*, and *KCTD15*), multiple LD blocks met selection criteria (genome-wide significance or replication). To

refine the 3-stage approach GRS, we asked whether the genotype for a single SNP could be used instead of the mean number of risk alleles at a locus. First, we identified the BMI-increasing allele for each SNP and calculated the linear association between the number of BMI-increasing alleles for that SNP and BMI measured at the first ARIC study visit. We next compared test-statistics and effect sizes between SNPs at each locus to identify the "lead-SNP", the SNP with the strongest association, and the worst-associated SNP. We then compared the effect size for the lead-SNP to the effect sizes for the worst-associated SNP and for the mean number of risk alleles across SNPs at the locus. These analyses asked 1) whether there was any difference in the signal from the different SNPs in a correlated set; and 2) whether a single SNP could provide an adequate summary of obesity-associated variation at the locus. Models were fitted using linear regression with statistical adjustment for demographics and geography. We compared effect sizes using the seemingly unrelated regression method (Baltagi 1980; Verzilli, Stallard, and Whittaker 2005). **Supplementary Table 6** shows results from this analysis. At all loci, the lead SNP, worst-associated SNP, and mean number of risk alleles performed similarly, with the exception of the *FTO* locus, at which the lead SNP rs9939609 performed slightly better than the worst-associated SNP rs1477196. Finally, we tested whether including multiple SNPs at a locus improved the prediction of BMI in a regression model. Analyses were conducted using the variable selection algorithm in the Stata program mfp (Royston and Ambler 1999). Details of this method are reported elsewhere (Royston and Sauerbrei 2003). Briefly, SNPs were added to a baseline model predicting BMI as a function of age, sex, and geography in order of decreasing statistical significance of the SNPs' bivariate association with BMI. SNPs were retained in the model if their inclusion resulted in a statistically significant ($p<0.05$) decrease in model deviance. Results showed that model fit was not improved by the inclusion of multiple SNPs at any locus. Therefore, we retained only the best-associated SNPs from each of the 7 loci, resulting in a 32-SNP GRS (**Supplementary Table 7**).

PART 3. SENSITIVITY ANALYSES TO TEST HETEROGENEITY IN GRS ASSOCIATIONS

We tested the linearity of GRS-BMI associations using quadratic and cubic specifications of the GRS in linear regression models. Coefficients for the higher order (i.e. squared and cubic) GRS terms were not statistically significant ($p>0.10$ for all), indicating that the GRS-BMI association was approximately linear. We tested the measurement specificity of GRS-BMI associations by comparing GRS effect sizes for BMI to GRS effect sizes for weight and for waist circumference using the seemingly unrelated regression method (Baltagi 1980). GRS coefficients were similar across all three models

(p>0.10 for tests of differences), indicating that the GRS predicted not just BMI, but related measures of body size and adiposity. We tested the whether GRS-BMI associations were different for men and women or for older as compared to younger individuals using product terms in linear regression models. Coefficients for product terms were not statistically significant (p>0.10 for all), indicating that GRS-BMI associations were similar for men and women and across early to late mid-life. Finally, we tested whether GRS-BMI associations differed across the 4 in-person assessments in the ARIC Study using the seemingly unrelated regression method. GRS effect sizes were similar across all 4 assessments (p>0.10 for all comparisons), indicating that GRS-BMI associations were consistent across measurement intervals.

Supplementary Table 1. Genome Wide Association Studies Included In 3-Stage Approach Analyses. GWAS information comes from the NHGRI GWAS Catalog (www.genome.gov). Risk SNPs were defined as any SNP associated with an obesity-related phenotype (BMI, weight, waist circumference, categorical obesity) at $p<10^{-5}$ in the discovery or combined discovery and replication samples of the GWAS. *Italicized counts include imputed genotypes; **Lindgren et al. also investigated associations with waist circumference, and these are the association tests included in the SNP selection analysis; ***Scherag et al. also investigated associations with BMI and both phenotypes were included in the SNP selection analysis. Citations for the GWAS are included as (Cotsapas et al. 2009; Fox et al. 2007; Frayling et al. 2007; Heard-Costa et al. 2009; Herbert et al. 2006; Hinney et al. 2007; Johansson et al. 2010; Lindgren et al. 2009; Liu et al. 2010; Liu et al. 2008; Loos et al. 2008; Meyre et al. 2009; Scherag et al. 2010; Scuteri et al. 2007; Speliotes et al. 2010; Thorleifsson et al. 2009; Willer et al. 2009).

	GWAS Chip Manufacturer	SNPs Genotyped*	SNPs in GWAS Catalog		Risk SNPs Included in Analyses
			SNPs	Phenotypes	
Herbert et al. 2006	Affymetrix	86,604	0	Obesity	0
Frayling et al. 2007	Affymetrix	490,032	1	BMI	1
Scuteri et al. 2007	Affymetrix	362,129	1	BMI, Weight	12
Fox et al. 2007	Affymetrix	70,897	5	BMI, Waist Circumference	12
Hinney et al. 2007	Affymetrix	440,794	1	Obesity (early onset extreme)	15
Liu et al. 2008	Affymetrix	379,319	0	Obesity	3
Loos et al. 2008	Affymetrix	344,883	2	BMI	10
Thorleifsson et al. 2009	Illumina	305,846	18	BMI, Weight	47
Willer et al. 2009	Affymetrix & Illumina	*2,399,588*	11	BMI	24
Meyre et al. 2009	Illumina	308,846	5	Obesity	32
Cotsapas et al. 2009	Illumina	457,251	13	Obesity (extreme)	15
Lindgren et al. 2009	Affymetrix & Illumina	2,573,738	NA	Adiposity**	10
Heard-Costa et al. 2009	Affymetrix & Illumina	512,349	7	Waist Circumference	320
Johansson et al. 2009	Illumina	318,237	17	BMI, Weight	26
Liu et al. 2010	Illumina	*559,712*	2	BMI	3
Scherag et al. 2010	Affymetrix & Illumina	*1,596,878*	2	Obesity (extreme)***	13
Speliotes et al. 2010	Affymetrix, Illumina, Perlegen	*~2.8 million*	38	BMI	42

Supplementary Table 2. Risk SNPs and Source Publications: All SNPs reported as associated with Obesity, BMI, Weight, or Waist Circumference at $p<1x10^{-5}$ in Discovery or Combined Discovery and Replication Samples

Risk SNP	Trait	Publication
rs9939609	BMI	Frayling et al. 2007 Science
rs1121980	BMI	
rs6602024	BMI	
rs7193144	BMI	
rs8050136	BMI	
rs9926289	BMI	
rs9930506	BMI	Scuteri et al. 2007
rs9939609	BMI	
rs9939973	BMI	
rs9940128	BMI	
*rs4512445**	Waist Circumference	
rs7193144	Waist Circumference	
rs8050136	Waist Circumference	
rs1106683	BMI	
rs1106684	BMI	
rs1333026	BMI	
rs10488165	Waist Circumference	
rs10504576	Waist Circumference	
rs1875517	Waist Circumference	Fox et al. 2007
rs2206682	Waist Circumference	
rs2223662	Waist Circumference	
rs4469448	Waist Circumference	
rs4471028	Waist Circumference	
rs6996971	Waist Circumference	
rs953536	Waist Circumference	
rs10008032	Extreme Obesity	
rs1121980	Extreme Obesity	
rs16998603	Extreme Obesity	
rs2172478	Extreme Obesity	
rs2969001	Extreme Obesity	
rs3783950	Extreme Obesity	
rs41492957	Extreme Obesity	
rs6076920	Extreme Obesity	Hinney et al. 2007
rs619819	Extreme Obesity	
rs7193144	Extreme Obesity	
rs8050136	Extreme Obesity	
rs9276431	Extreme Obesity	
rs9939609	Extreme Obesity	
rs9939973	Extreme Obesity	
rs9940128	Extreme Obesity	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs16986921	BMI	
rs6013029	BMI	
rs6020712	BMI	Liu et al. 2008
rs10498767	BMI	
rs1121980	BMI	
rs17700633	BMI	
rs17782313	BMI	
rs2572106	BMI	Loos et al. 2008
rs2679120	BMI	
rs4623795	BMI	
rs7212681	BMI	
rs7336049	BMI	
rs748192	BMI	
rs10501087	BMI	
rs10783050	BMI	
rs10913469	BMI	
rs12970134	BMI	
rs1776012	BMI	
rs2568958	BMI	
rs2867125	BMI	
rs29941	BMI	
rs3101336	BMI	
rs3751812	BMI	
rs4074134	BMI	
rs467650	BMI	
rs4788102	BMI	
rs4854344	BMI	
rs4923461	BMI	
rs6265	BMI	
rs6499640	BMI	
rs7138803	BMI	
rs7190492	BMI	
rs7336332	BMI	
rs7481311	BMI	
rs7498665	BMI	
rs7561317	BMI	Thorleifsson et al. 2009
rs7647305	BMI	
rs7647305	BMI	
rs8044769	BMI	
rs8049439	BMI	
rs8050136	BMI	
rs836964	BMI	
rs867559	BMI	
rs925946	BMI	
rs9424977	BMI	
rs1047440	Weight	
rs1077393	Weight	
rs10835211	Weight	
rs1350341	Weight	
rs1350341	Weight	
rs17069257	Weight	
rs1973993	Weight	
rs2115172	Weight	
rs2260000	Weight	
rs2260000	Weight	
rs2844479	Weight	
rs2844479	Weight	
rs3766431	Weight	
rs633265	Weight	
rs6477693	Weight	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs10769908	BMI	
rs10769908	BMI	
rs10838738	BMI	
rs10838738	BMI	
rs10938397	BMI	
rs10938397	BMI	
rs11084753	BMI	
rs11084753	BMI	
rs11084753	BMI	
rs11773921	BMI	
rs12324805	BMI	
rs1421085	BMI	
rs1439845	BMI	
rs17700144	BMI	
rs17782313	BMI	
rs17782313	BMI	
rs2145270	BMI	Willer et al. 2009
rs2145270	BMI	
rs2245715	BMI	
rs2815752	BMI	
rs2815752	BMI	
rs2815752	BMI	
rs4752856	BMI	
rs6548238	BMI	
rs6548238	BMI	
rs6548238	BMI	
rs6907460	BMI	
rs7181095	BMI	
rs7498665	BMI	
rs7498665	BMI	
rs752238	BMI	
rs9931989	BMI	
rs9939609	BMI	
rs9939609	BMI	
rs10508503	Obesity	
rs11071927	Obesity	
rs11956401	Obesity	
rs12588659	Obesity	
rs12633433	Obesity	
rs1326986	Obesity	
rs1343772	Obesity	
rs1380100	Obesity	
rs1396618	Obesity	
rs1421085	Obesity	
rs1424233	Obesity	
rs16829231	Obesity	
rs17782313	Obesity	
rs1805081	Obesity	
rs1858367	Obesity	
rs2011946	Obesity	
rs2158044	Obesity	Meyere et al. 2009
rs2908338	Obesity	
rs3026762	Obesity	
rs3102841	Obesity	
rs413693	Obesity	
rs4712652	Obesity	
rs4786847	Obesity	
rs6463923	Obesity	
rs646839	Obesity	
rs6580742	Obesity	
rs6796959	Obesity	
rs7506051	Obesity	
rs7717673	Obesity	
rs908078	Obesity	
rs9275582	Obesity	
rs987052	Obesity	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs10433903	Extreme Obesity	
rs10999409	Extreme Obesity	
rs12295638	Extreme Obesity	
rs12492816	Extreme Obesity	
rs12635698	Extreme Obesity	
rs1435703	Extreme Obesity	
rs2274459	Extreme Obesity	
rs374748	Extreme Obesity	Cotsapas et al. 2009
rs6110577	Extreme Obesity	
rs6726292	Extreme Obesity	
rs7474896	Extreme Obesity	
rs7603514	Extreme Obesity	
rs9366829	Extreme Obesity	
rs9941349	Extreme Obesity	
rs999943	Extreme Obesity	
rs10085177	Waist Circumference	
rs11970116	Waist Circumference	
rs13116494	Waist Circumference	
rs2245667	Waist Circumference	
rs4737325	Waist Circumference	Lindgren et al. 2009
rs6429082	Waist Circumference	
rs7194591	Waist Circumference	
rs7826222	Waist Circumference	
rs7970350	Waist Circumference	
rs987237	Waist Circumference	
rs10096750	BMI	
rs10145154	BMI	
rs10146997	BMI	
rs10150332	BMI	
rs10173167	BMI	
rs10188334	BMI	
rs10189761	BMI	
rs10190052	BMI	
rs10193244	BMI	
rs10511835	BMI	
rs10813208	BMI	
rs10852521	BMI	
rs10871777	BMI	
rs10875982	BMI	
rs10969478	BMI	
rs11075985	BMI	
rs11075987	BMI	
rs11075989	BMI	
rs11075990	BMI	
rs11127483	BMI	
rs11127484	BMI	
rs11127485	BMI	Heard-Costa et al.
rs11127491	BMI	2009
rs11152213	BMI	
rs11169176	BMI	
rs1121980	BMI	
rs11520442	BMI	
rs11642841	BMI	
rs11660783	BMI	
rs11662368	BMI	
rs11663816	BMI	
rs11664883	BMI	
rs11665563	BMI	
rs12002080	BMI	
rs12149832	BMI	
rs12446228	BMI	
rs12623218	BMI	
rs12714414	BMI	
rs12714415	BMI	
rs12954782	BMI	
rs12955983	BMI	
rs12957347	BMI	
rs12960928	BMI	
rs12964203	BMI	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs12966550	BMI	
rs12967135	BMI	
rs12969709	BMI	
rs12970134	BMI	
rs12992154	BMI	
rs12995480	BMI	
rs13007080	BMI	
rs13007086	BMI	
rs13012571	BMI	
rs13021737	BMI	
rs1320330	BMI	
rs1320331	BMI	
rs1320336	BMI	
rs1320337	BMI	
rs1320338	BMI	
rs13386517	BMI	
rs13386627	BMI	
rs13386964	BMI	
rs13388043	BMI	
rs13393304	BMI	
rs13396935	BMI	
rs13397165	BMI	
rs13401686	BMI	
rs13415094	BMI	
rs1350341	BMI	
rs1421085	BMI	
rs1456404	BMI	
rs1457489	BMI	
rs1477196	BMI	
rs1539952	BMI	
rs1553754	BMI	
rs1555967	BMI	
rs1558902	BMI	
rs1619975	BMI	
rs1673518	BMI	Heard-Costa et al. 2009
rs17109256	BMI	
rs17175643	BMI	
rs17201502	BMI	
rs17299673	BMI	
rs17700144	BMI	
rs17782313	BMI	
rs17817288	BMI	
rs17817449	BMI	
rs17817964	BMI	
rs1861866	BMI	
rs1861867	BMI	
rs1942860	BMI	
rs1942863	BMI	
rs1942866	BMI	
rs2051311	BMI	
rs2051312	BMI	
rs2058908	BMI	
rs2168708	BMI	
rs2168711	BMI	
rs2206277	BMI	
rs2288278	BMI	
rs2331841	BMI	
rs2397026	BMI	
rs2860323	BMI	
rs2867108	BMI	
rs2867109	BMI	
rs2867110	BMI	
rs2867112	BMI	
rs2867113	BMI	
rs2867122	BMI	
rs2867123	BMI	
rs2867125	BMI	
rs2867131	BMI	
rs2903492	BMI	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs2947411	BMI	
rs297924	BMI	
rs34341	BMI	
rs3751812	BMI	
rs3751813	BMI	
rs3928247	BMI	
rs4045166	BMI	
rs4299252	BMI	
rs4423631	BMI	
rs4438957	BMI	
rs4452188	BMI	
rs4613321	BMI	
rs4615388	BMI	
rs4620360	BMI	
rs474112	BMI	
rs475134	BMI	
rs476828	BMI	
rs4783819	BMI	
rs4784323	BMI	
rs4793927	BMI	
rs4854344	BMI	
rs4854348	BMI	
rs4854349	BMI	
rs487720	BMI	
rs489693	BMI	
rs492443	BMI	
rs497353	BMI	
rs5017300	BMI	
rs5017303	BMI	
rs521663	BMI	
rs523288	BMI	
rs536783	BMI	
rs538656	BMI	
rs545708	BMI	
rs559623	BMI	Heard-Costa et al. 2009
rs562622	BMI	
rs563726	BMI	
rs565239	BMI	
rs565970	BMI	
rs571312	BMI	
rs574988	BMI	
rs589850	BMI	
rs590215	BMI	
rs591166	BMI	
rs611428	BMI	
rs633265	BMI	
rs649721	BMI	
rs6499640	BMI	
rs6548237	BMI	
rs6567155	BMI	
rs6567160	BMI	
rs6567161	BMI	
rs663129	BMI	
rs666181	BMI	
rs6711012	BMI	
rs6719518	BMI	
rs6719980	BMI	
rs6725549	BMI	
rs6728726	BMI	
rs6731348	BMI	
rs6731688	BMI	
rs6732471	BMI	
rs6734363	BMI	
rs6742576	BMI	
rs6743060	BMI	
rs6744646	BMI	
rs6744653	BMI	
rs6745266	BMI	
rs6752470	BMI	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs6755502	BMI	
rs681630	BMI	
rs682614	BMI	
rs683430	BMI	
rs7022642	BMI	
rs7132908	BMI	
rs7138803	BMI	
rs7144011	BMI	
rs7185735	BMI	
rs7190492	BMI	
rs7193144	BMI	
rs7201850	BMI	
rs7202116	BMI	
rs7203521	BMI	
rs7205986	BMI	
rs7206010	BMI	
rs7206790	BMI	
rs7240566	BMI	
rs7338657	BMI	
rs7561317	BMI	
rs7567570	BMI	
rs7570198	BMI	
rs7571957	BMI	
rs7574359	BMI	
rs7576624	BMI	
rs7576635	BMI	
rs7585056	BMI	
rs7587786	BMI	
rs7604609	BMI	
rs7608050	BMI	
rs7715806	BMI	
rs7831920	BMI	
rs8043757	BMI	Heard-Costa et al.
rs8044769	BMI	2009
rs8047395	BMI	
rs8050136	BMI	
rs8051591	BMI	
rs8055197	BMI	
rs8057044	BMI	
rs8083289	BMI	
rs8086627	BMI	
rs8089364	BMI	
rs8091524	BMI	
rs8095404	BMI	
rs921971	BMI	
rs939582	BMI	
rs939583	BMI	
rs953442	BMI	
rs975918	BMI	
rs981106	BMI	
rs981113	BMI	
rs987237	BMI	
rs9922047	BMI	
rs9922619	BMI	
rs9922708	BMI	
rs9923147	BMI	
rs9923233	BMI	
rs9923544	BMI	
rs9928094	BMI	
rs9930333	BMI	
rs9930501	BMI	
rs9930506	BMI	
rs9931494	BMI	
rs9932754	BMI	
rs9935401	BMI	
rs9936385	BMI	
rs9937053	BMI	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs993887	BMI	
rs9939609	BMI	
rs9939973	BMI	
rs9940128	BMI	
rs9940646	BMI	
rs9941349	BMI	
rs10059683	Waist Circumference	
rs10066756	Waist Circumference	
rs10068332	Waist Circumference	
rs10146690	Waist Circumference	
rs10150482	Waist Circumference	
rs10869557	Waist Circumference	
rs10869558	Waist Circumference	
rs10869559	Waist Circumference	
rs11778132	Waist Circumference	
rs11780082	Waist Circumference	
rs11857639	Waist Circumference	
rs11990688	Waist Circumference	
rs12271537	Waist Circumference	
rs12274672	Waist Circumference	
rs12475139	Waist Circumference	
rs12792768	Waist Circumference	
rs13404551	Waist Circumference	
rs1447905	Waist Circumference	
rs1521252	Waist Circumference	
rs16930931	Waist Circumference	
rs17008958	Waist Circumference	
rs17061143	Waist Circumference	
rs17109221	Waist Circumference	
rs17476669	Waist Circumference	
rs17537900	Waist Circumference	
rs17836088	Waist Circumference	
rs2164210	Waist Circumference	
rs2236783	Waist Circumference	
rs2322659	Waist Circumference	Heard-Costa et al.
rs2322660	Waist Circumference	2009
rs2365642	Waist Circumference	
rs2370982	Waist Circumference	
rs303211	Waist Circumference	
rs309134	Waist Circumference	
rs309137	Waist Circumference	
rs309160	Waist Circumference	
rs309168	Waist Circumference	
rs4098360	Waist Circumference	
rs4420638	Waist Circumference	
rs4701252	Waist Circumference	
rs4758213	Waist Circumference	
rs4758215	Waist Circumference	
rs507824	Waist Circumference	
rs569406	Waist Circumference	
rs6499641	Waist Circumference	
rs6714750	Waist Circumference	
rs6716536	Waist Circumference	
rs6754311	Waist Circumference	
rs6817633	Waist Circumference	
rs6837818	Waist Circumference	
rs6870971	Waist Circumference	
rs687670	Waist Circumference	
rs693895	Waist Circumference	
rs6998794	Waist Circumference	
rs7110070	Waist Circumference	
rs7156625	Waist Circumference	
rs745500	Waist Circumference	
rs748841	Waist Circumference	
rs7579771	Waist Circumference	
rs7824886	Waist Circumference	
rs7932813	Waist Circumference	
rs8059991	Waist Circumference	
rs892715	Waist Circumference	
rs9598518	Waist Circumference	
rs9790104	Waist Circumference	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs1024889	BMI	
rs1152846	BMI	
rs12517906	BMI	
rs1458095	BMI	
rs1878047	BMI	
rs1927702	BMI	
rs2383393	BMI	
rs3803915	BMI	
rs3803915	BMI	
rs3934834	BMI	
rs4085400	BMI	
rs824931	BMI	
rs875283	BMI	Johansson et al. 2009
rs10844154	Weight	
rs10972341	Weight	
rs10972350	Weight	
rs1152846	Weight	
rs12517906	Weight	
rs1570885	Weight	
rs1816002	Weight	
rs1840440	Weight	
rs2765086	Weight	
rs4879869	Weight	
rs7209395	Weight	
rs7919006	Weight	
rs965178	Weight	
rs2275215	BMI	Liu et al. 2010
rs10458787	BMI	
rs11127485	BMI	
rs1558902	BMI	
rs9935401	BMI	
rs10926984	Obesity	
rs12145833	Obesity	
rs2783963	Obesity	
rs11127485	Obesity	Scherag et al. 2010
rs17150703	Obesity	
rs13278851	Obesity	
rs516175	Obesity	
rs1558902	Obesity	
rs9935401	Obesity	
rs17700144**	Obesity	

Supplementary Table 2 Continued

Risk SNP	Trait	Publication
rs1558902	BMI	
rs2860323	BMI	
rs6567160	BMI	
rs10938397	BMI	
rs10767664	BMI	
rs543874	BMI	
rs2815752	BMI	
rs10182181	BMI	
rs12444979	BMI	
rs7498665	BMI	
rs987237	BMI	
rs2241423	BMI	
rs9816226	BMI	
rs7138803	BMI	
rs2287019	BMI	
rs1514177	BMI	
rs13107325	BMI	
rs2112347	BMI	
rs10968576	BMI	
rs3817334	BMI	
rs3810291	BMI	Speliotes et al. 2010
rs887912	BMI	
rs10150332	BMI	
rs7640855	BMI	
rs11847697	BMI	
rs2890652	BMI	
rs11165643	BMI	
rs4771122	BMI	
rs4836133	BMI	
rs4929949	BMI	
rs29938	BMI	
rs9296115	BMI	
rs2922763	BMI	
rs2444217	BMI	
rs867559	BMI	
rs3764400	BMI	
rs255414	BMI	
rs6955651	BMI	
rs17016663	BMI	
rs6477694	BMI	
rs2652594	BMI	
rs2035935	BMI	

Supplementary Table 2 Footnote: *Reported as "SNP_A-2284869" and crosswalked to rs ID using the Affy 6.0 SNP name to rs ID crosswalk file "GenomeWideSNP_6.na30.annot.csv"; **The GWAS catalog reports rs10871777 (in LD with rs17700144 at R^2=0.85) as the obesity-associated SNP near the gene MC4R in Scherag et al. SNPs are reported only once per GWAS. Associations are reported for BMI where present and for other phenotypes where BMI was not investigated or the SNP was not associated with BMI at $p<1 \times 10^{-5}$

Supplementary Table 3. Replicated and/or Genome-Wide Significant LD Blocks Identified in 3-Stage Approach Analyses. LD blocks were defined from LD analyses of risk SNPs (genotype-phenotype association at $p<1x10^{-5}$) using data from the HapMap version 3 CEU sample accessed via Seattle SNPs's Genome Variation Server and an LD threshold of $R^2 \geq 0.95$. Replication was evaluated as the number of GWAS reporting any SNP in the block as a risk SNP. Genes were evaluated within 100kb in either direction from an LD block's outermost SNPs.

Chromsome	Identified LD Blocks	Replicated LD Blocks	Mean Number of Replications (All Blocks)	Genes
1	4	3	2.0	*NEGR1, TNNI3K, PTB2, SEC16B*
2	6	2	2.0	*LRP1B, TMEM18*
3	3	0	1.0	*CADM2, ETV5/DGKG*
4	2	1	1.5	*GNPDA2, SLC39A8*
5	2	0	1.0	*POC5, ZNF608*
6	1	1	3.0	*TFAP2B*
9	2	1	1.5	*LING02/LRRN6C, LMX1B*
11	7	0	1.0	*RPL27A, BDNF, MTCH2*
12	1	1	3.0	*BDCDIN3D/FAIM2/NCKAP5L*
13	1	0	3.0	*MTIF3, GRF3A*
14	2	1	1.5	*PRKD1, NRXN3*
15	1	0	1.0	*MAP2K5*
16	26	14	3.0	*GRP5B, ATXN2L/TUFM/SH2B1, FTO*
18	7	7	2.6	*MC4R*
19	4	1	1.3	*KCTD15, ZC3H4, QPCTL, TMEM160*

Supplementary Table 4. Characteristics of Replicated and/or Genome-Wide Significant LD Blocks

			LD Block		GWAS Publication															
Chrom-osome	Chromosomal Space Covered by All Risk SNPs in the LD Block (NCBI Build 36)	Nearby Genes	Seed SNPs (risk SNPs in LD with all risk SNPs in block at $R^2 \geq 0.95$) // Proxy SNPs (risk SNPs in LD with any seed SNP at R2≥0.95)	Any SNP in Block Genome-Wide Significant	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]
1	72,523,773 - 72,585,028	NEGR1	rs2568958, rs2815752, rs3101336	Yes							X	X								X
	74,763,990	TNNI3K	rs1514177	Yes																X
	96,696,685 - 96,716,582	*PTBP2*	rs11165643 // rs1973993	Yes							X									X
	176,156,103 - 176,180,142	SEC16B	rs10913469, rs543874	Yes							X									X
2	604,168 - 643,874	TMEM18	See footnote	Yes							X					X			X	X
	604,210 - 643,874	TMEM18		Yes							X					X			X	X
	624,905	TMEM18	rs6548238	Yes								X								
	25,003,800		rs10182181	Yes																X
	59,156,381		rs887912	Yes																X
	142,676,401	LRP1B	rs2890652	Yes																X
3	85,956,854	CADM2	rs7640855	Yes																X
	187,316,984	ETV5/DGKG	rs7647305	Yes							X									
	187,317,193	ETV5/DGKG	rs9816226	Yes																X
4	44,877,284		rs10938397	Yes								X								X
	103,407,732	SLC39A8	rs13107325	Yes																X
5	75,050,998	POC5	rs2112347	Yes																X
	124,360,002		rs4836133	Yes																X
6	50,906,485 - 50,911,009	TFAP2B	rs2206277, rs987237	Yes											X	X				X
9	28,404,339	LING02	rs10968576	Yes																X
	128,505,146	LMX1B	rs867559	$p<1x10^{-6}$							X									X
11	8,561,169	STK33	rs4929949	Yes																X
	27,603,861 - 27,626,684	BDNF	rs10501087, rs4074134, rs4923461	Yes							X									
	27,636,492	BDNF	rs6265	Yes							X									
	27,682,562	BDNF	rs10767664	Yes																X
	27,623,778 - 27,623,778	BDNF	rs925946	Yes							X									
	47,604,618 - 47,619,625	MTCH2	rs10838738, rs4752856	Yes								X								
	47,607,569	MTCH2	rs3817334	Yes																X
12	48,533,735	BDCDIN3D, FAIM2, NCKAP5L	rs7138803	Yes							X					X				X
13	26,918,180	MTIF3, *GRF3A*	rs4771122	Yes																X
14	29,584,863		rs11847697	Yes																X
	78,961,635 - 79,014,915	NRXN3	rs10145154, rs10150332, rs17109256, rs7144011 // rs10146997, rs10150482, rs17109221, rs17836088, rs7156625	Yes												X				X
15	65,873,892	MAP2K5	rs2241423	Yes																X

Supplementary Table 4 Continued

	LD Block				GWAS Publication														
Chrom-osome	Chromosomal Space Covered by All Risk SNPs in the LD Block (NCBI Build 36)	Genes Overlapping LD Block/ 10kb of SNP*	Seed SNPs (risk SNPs in LD with all risk SNPs in block at R^2≥0.95) // Proxy SNPs (risk SNPs in LD with any seed SNP at R2≥0.95)	Any SNP in Block Genome-Wide Significant	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]
	19,841,101	GPRC5B	rs12444979	Yes															X
	28,745,016 - 28,790,742	ATXN2L, TUFM, SH2B1	rs4788102, rs7498665, rs8049439	Yes							X	X							X
	52,312,678 - 52,327,178	FTO	rs6499640, rs7203521, rs7206010	Yes							X					X			
	52,355,409	FTO	rs7206790	Yes												X			
	52,356,024 - 52,361,841	FTO	rs8047395 //rs1861866, rs8055197	Yes												X			
	52,356,024 - 52,363,781	FTO	rs1861866, rs8055197 // rs10852521, rs8047395, rs9922047	Yes												X			
	52,360,657 - 52,372,662	FTO	rs10852521, rs9922047 // rs11075987, rs1861866, rs8055197	Yes												X			
	52,362,466 - 52,372,662	FTO	rs11075987 // rs10852521, rs9922047	Yes												X			
	52,365,265	FTO	rs17817288	Yes												X			
	52,370,115	FTO	rs8057044	Yes												X			
	52,396,636	FTO	rs8044769	Yes							X					X			
	52,357,008 - 52,366,748	FTO	rs11075985, rs9940646 // rs1121980, rs9923147, rs9923544, rs9928094, rs9930333, rs9937053, rs9939973, rs9940128	Yes		X		X		X						X			
	52,357,008 - 52,384,680	FTO	rs1121980, rs9923147, rs9923544, rs9928094, rs9930333, rs9937053, rs9939973, rs9940128 // rs11075985, rs1421085, rs1558902, rs7201850, rs9931494, rs9940646, rs9941349	Yes		X		X		X		X	X	X		X		X	X
	52,357,008 - 52,385,567	FTO	rs1421085, rs1558902 // rs17817964, rs7185735, rs7193144, rs7202116, rs9937053	Yes		X		X				X	X			X		X	X
16	52,357,008 - 52,389,272	FTO	rs7201850, rs9931494, rs9941349 // rs1121980, rs9922619, rs9922708, rs9923147, rs9923544, rs9928094, rs9930333, rs9930501, rs9930506, rs9932754, rs9937053, rs9939973, rs9940128	Yes		X		X		X				X		X			
	52,358,455 - 52,400,409	FTO	rs17817964, rs7185735 // rs11075989, rs11075990, rs12149832, rs1421085, rs1558902, rs17817449, rs3751812, rs7193144, rs7202116, rs8043757, rs8050136, rs8051591, rs9923233, rs9935401, rs9939609	Yes	X	X		X			X	X	X			X		X	X
	52,361,075 - 52,400,409	FTO	rs7193144, rs7202116 // rs11075989, rs11075990, rs12149832, rs1558902, rs17817449, rs17817964, rs3751812, rs7185735, rs8043757, rs8050136, rs8051591, rs9923233, rs9935401, rs9939609	Yes	X	X		X			X	X				X		X	X
	52,368,187 - 52,385,567	FTO	rs11075989, rs11075990, rs17817449, rs3751812, rs8043757, rs8050136, rs8051591, rs9923233, rs9935401, rs9939609 // rs17817964, rs7185735, rs7193144, rs7202116, rs9936385	Yes	X	X		X			X	X				X		X	
	52,368,187 - 52,400,409	FTO	rs12149832 // rs17817964, rs7185735, rs7193144, rs7202116	Yes		X		X								X			
	52,376,670 - 52,377,378	FTO	rs9936385 // rs11075989, rs9923233	Yes												X			
	52,379,363 - 52,389,272	FTO	rs9922619, rs9922708, rs9930501, rs9932754 // rs7201850, rs9930506, rs9931494, rs9941349	Yes		X								X		X			
	52,382,989 - 52,389,272	FTO	rs9930506 // rs9922619, rs9922708, rs9930501, rs9931494, rs9932754, rs9941349	Yes		X								X		X			
	52,406,062	FTO	rs1861867	Yes												X			
	52,357,888 - 52,386,253	FTO	rs12446228, rs1477196, rs4783819, rs7190492	Yes							X					X			
	52,376,209	FTO	rs3751813	Yes												X			
	52,402,988	FTO	rs11642841	Yes												X			

Supplementary Table 4 Continued

	LD Block				GWAS Publication														
Chrom-osome	Chromosomal Space Covered by All Risk SNPs in the LD Block (NCBI Build 36)	Genes Overlapping LD Block/ 10kb of SNP*	Seed SNPs (risk SNPs in LD with all risk SNPs in block at R^2≥0.95) // Proxy SNPs (risk SNPs in LD with any seed SNP at R2≥0.95)	Any SNP in Block Genome-Wide Significant	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]
18	55,962,962	*MC4R*	rs17700144	$p<1x10^{-6}$								X				X			X	
	55,980,115 - 56,003,928	*MC4R*	rs10871777, rs11152213, rs12967135, rs17782313, rs2168711, rs476828, rs523288, rs538656, rs571312, rs6567160, rs663129	Yes						X		X	X			X				X
	55,964,628 - 56,003,732	*MC4R*	rs1350341, rs1619975, rs1673518, rs2051311, rs2051312, rs2331841, rs474112, rs475134, rs487720, rs536783, rs545708, rs559623, rs562622, rs565239, rs565970, rs574988, rs589850, rs591166, rs611428, rs649721, rs6567161, rs666181, rs681630, rs682614, rs683430, rs975918, rs993887 // rs521663, rs633265	$p<1x10^{-6}$								X				X				
	56,009,782 - 56,048,783	*MC4R*	rs12960928 // rs11663816, rs11664883, rs11665563, rs12954782, rs12969709, rs12970134, rs1457489, rs17175643, rs492443, rs8083289, rs8089364, rs921971	Yes								X				X				
	56,009,782 - 56,062,310	*MC4R*	rs921971 // rs11663816, rs11664883, rs11665563, rs12954782, rs12955983, rs12960928, rs12964203, rs12966550, rs12969709, rs12970134, rs1457489, rs17175643, rs2168708, rs492443, rs8083289, rs8089364	Yes								X				X				
	56,009,809 - 56,047,722	*MC4R*	rs12955983 // rs11663816, rs11664883, rs11665563, rs12954782, rs12969709, rs12970134, rs1457489, rs17175643, rs8083289, rs8089364, rs921971	Yes								X				X				
	56,009,809 - 56,062,310	*MC4R*	rs11663816, rs11664883, rs11665563, rs12954782, rs12964203, rs12966550, rs12969709, rs12970134, rs1457489, rs17175643, rs2168708, rs8083289, rs8089364 // rs12955983, rs12960928, rs921971	Yes								X				X				
19	39,001,372 - 39,003,321	KCTD15	rs29938, rs29941	Yes								X								X
	39,013,977	*KCTD15*	rs11084753	Yes									X							
	52,260,843	ZC3H4, TMEM160	rs3810291	Yes																X
	50,894,012	QPCTL	rs2287019	Yes																X

Supplementary Table 4 Footnote: GWAS are numbered as follows: [1] Frayling et al. 2007, *Science*; [2] Scuteri et al. 2007, *PLoS Genetics*; [3] Fox et al. 2007, *BMC Medical Genetics*; [4] Hinney et al. 2007, *PLoS One*; [5] Liu et al. 2008, *Human Molecular Genetics*; [6] Loos et al. 2008, *Nature Genetics*; [7] Thorleifsson et al. 2009, *Nature Genetics*; [8] Willer et al. 2009, *Nature Genetics*; [9] Meyere et al. 2009 *Nature Genetics*; [10] Cotsapas et al. 2009, *Human Molecular Genetics*; [11] Lindgren et al. 2009 *PLoS Genetics*; [12] Heard-Costa et al. 2009, *PLoS Genetics*; [13] Johansson et al. 2009, *Obesity*; [14] Liu et al. 2010, *Twin Research and Human Genetics*; [15] Shcerag et al. 2010, *PLoS Genetics*; Speliotes et al. 2010, *Nature Genetics*. LD Blocks were defined using an R^2 threshold of 0.95. Genes are reported within 100 kb of any seed SNP. Italicized genes fall outside the 100kb range, but contain SNPs in LD with a block seed. GWAS are indicated as replicating a block if they reported a SNP in LD at R^2≥0.95 with a block seed or proxy as associated with an obesity-related phenotype at $p<1x10^{-5}$ in either their discovery or combined discovery and replication samples.

Block 2.2: (**seeds**) rs10173167, rs10188334, rs10189761, rs10190052, rs10193244, rs11127484, rs11127485, rs11127491, rs12714414, rs12714415, rs12992154, rs12995480, rs13007080, rs13007086, rs13012571, rs13021737, rs1320331, rs1320336, rs1320337, rs1320338, rs13386517, rs13386627, rs13386964, rs13388043, rs13393304, rs13396935, rs13397165, rs13401686, rs13415094, rs2860323, rs2867108, rs2867109, rs2867110, rs2867112, rs2867113, rs2867122, rs2867125, rs2903492, rs2947411, rs4423631, rs4452188, rs4613321, rs4854344, rs4854348, rs4854349, rs5017300, rs5017303, rs6711012, rs6719518, rs6719980, rs6725549, rs6728726, rs6731348, rs6731688,

rs6732471, rs6734363, rs6743060, rs6744646, rs6744653, rs6752470, rs6755502, rs7561317, rs7567570, rs7570198, rs7571957, rs7574359, rs7576624, rs7576635, rs7585056, rs7604609, rs7608050, rs939582, rs939583

Block 2.3: (**seeds**) rs2867123, (**proxies**) rs10173167, rs10188334, rs10189761, rs10190052, rs10193244, rs11127484, rs11127485, rs11127491, rs12714414, rs12714415, rs12992154, rs12995480, rs13007080, rs13007086, rs13012571, rs13021737, rs1320331, rs1320336, rs1320337, rs1320338, rs13386517, rs13386627, rs13386964, rs13388043, rs13393304, rs13396935, rs13397165, rs13401686, rs13415094, rs2860323, rs2867108, rs2867109, rs2867110, rs2867112, rs2867113, rs2867122, rs2867123, rs2867125, rs2903492, rs4423631, rs4452188, rs4613321, rs4854344, rs4854348, rs4854349, rs5017300, rs5017303, rs6711012, rs6719518, rs6719980, rs6725549, rs6728726, rs6731348, rs6731688, rs6732471, rs6734363, rs6743060, rs6744646, rs6744653, rs6752470, rs6755502, rs7561317, rs7567570, rs7570198, rs7571957, rs7574359, rs7576624, rs7576635, rs7585056, rs7604609, rs7608050, rs939582, rs939583

Supplementary Table 5. Effect Sizes for Genetic Risk Scores Created Using the 3-Stage Approach and the Best-Guess and Top-Hits Approaches. To measure BMI effect sizes for the GRSs, we estimated Pearson correlations (r) from separate linear regressions of BMI on each of the GRSs. To measure obesity effect sizes for the GRSs, we estimated odds ratios (OR) from separate logistic regressions of obesity on each of the GRSs. Regression models were adjusted for age (linear and quadratic terms), gender, the age-gender interaction, and the ARIC Study Centers where data were collected. In Panel A, the Best-Guess GRS was based on the GRS published by Li and colleagues (Li et al. 2010) and the Top-Hits GRS was based on the GRS published by Peterson and colleagues (Peterson et al. 2011). In Panel B, the Best Guess GRS was based on the full set of obesity- and BMI-associated SNPs listed in the NHGRI GWAS Catalog and the Top-Hits GRS was based on the GRS published by Speliotes and colleagues (Speliotes et al. 2010). ***p<0.001. Comparison of effect sizes using the seemingly unrelated regression method (Baltagi 1980) indicated that effect sizes for the 3 GRSs in Panel A were not statistically different from one another (p-value for difference >0.10 for all), but that among the GRSs in Panel B, the 3-stage approach performed better than the Best-Guess and Top-Hits GRSs (p<0.05 for all). However, our sample had only 40% power to detect effect size differences of r=0.01 / OR=1.01, so this result should be interpreted with caution.

		Effect Sizes	
		BMI	Obesity
Approach to GRS Construction	SNPs	Pearson Correlation (r)	Odds Ratio [95% CI]
Panel A. GRSs Constructed from Results of 9 GWAS Published by December 31, 2008			
3-Stage	28	0.08***	1.08 [1.06-1.10]
Best-Guess	12	0.08***	1.08 [1.06-1.11]
Top-Hits	59	0.06***	1.07 [1.04-1.09]
Panel B. GRSs Constructed from Results of the Full Set of 16 GWAS			
3-Stage	57	0.11***	1.12 [1.10-1.15]
Best-Guess	97	0.10***	1.11 [1.09-1.13]
Top-Hits	32	0.10***	1.10 [1.08-1.12]

Supplementary Table 6. Analysis of Loci with Multiple Tag SNPs. * "Lead SNP" is underlined; "Worst-associated SNP" is italicized; Test statistics and effect sizes were estimated in linear regression models of BMI adjusted for demographics and geography. "Lead SNPs" and "Worst-associated SNPs" were determined from the test statistics for the individual SNPs. Effect sizes were compared using the seemingly unrelated regressions method (Baltagi 1980).

Locus	ARIC SNPs Tagging LD Blocks in Genic Region	Minimum R^2 Among Tag SNPs	Effect Size (Pearson's r) p-value for comparison with lead SNP		
			Lead SNP	Worst-Associated SNP	Mean Number of BMI-Increasing Alleles
Chr 2 *TMEM18*	*rs10189761* , rs2867123, rs4854345	0.94	0.027	0.023 p=0.276	0.025 p=0.371
Chr 3 ETV5/DGKG	rs12516728, *rs9863591*	0.85	0.007	<0.001 p=0.721	0.018 p=0.427
Chr 11 BDNF	rs10501087, *rs7103411* , rs6265, rs11030108	0.86	0.027	0.022 p=0.124	0.026 p=0.485
Chr 11 MTCH	rs12419692, *rs3817334*	0.77	0.020	0.019 p=0.871	0.020 p=0.878
Chr 16 FTO	*rs1477196* , rs17817288, rs1121980, rs9922047, rs9939973, rs9940128, rs9941349, rs7193144, rs7203521, rs9939609, rs8050136, rs9930506	0.40	0.072	0.034 p<0.001	0.068 p=0.104
Chr 18 MC4R	*rs476828* , rs1673518, rs17782313, rs11663816, rs11665563, rs12969709, rs12970134	0.25	0.026	0.019 p=0.158	0.025 p=0.062
Chr 19 KCDT15	*rs29942* , rs11084753	0.58	0.010	0.009 p=0.879	0.009 p=0.913

Supplementary Table 7. SNPs Included in the Obesity Genetic Risk Score.

								White Participants, n=8,210-8,8,286				Black Participants, n=2,402-2,442			
Chr	Nearby Gene	Tag SNP	GWAS Replications	BMI-Increasing Allele in GWAS	Test Alelle	Other Allele	Effect-Size Weight	Test Allele Frequency	Per Allele Change in BMI	p-value	Direction of Association Inconsistent with GWAS	Test Allele Frequency	Per Allele Change in BMI	p-value	Direction of Association Inconsistent with GWAS
1	NEGR1	rs2815752	3	Major	G	A	0.13	38%	-0.259	0.001		45%	-0.071	0.673	
	TNNI3K	rs1514175	1	Minor	A	G	0.07	43%	-0.001	0.985		68%	-0.091	0.608	X
	PTBP2	rs1555543	2	Major	A	C	0.06	42%	-0.128	0.086		57%	-0.031	0.855	
	SEC16B	rs543874	2	Minor	G	A	0.22	20%	0.341	0.000		25%	0.335	0.095	
2	FANCL	rs759250	1	Minor	A	G	0.10	29%	0.036	0.656		8%	-0.242	0.475	X
	LRP1B	rs2121279	1	Minor	T	C	0.08	14%	0.234	0.032		3%	-0.253	0.651	X
	TMEM18	rs2867123	5	Major	G	C	0.30	17%	-0.237	0.018		12%	0.022	0.935	X
	RBJ	rs10182181	1	Minor	G	A	0.14	46%	0.117	0.117		84%	0.758	0.001	
3	CADM2	rs12714640	1	Minor	A	C	0.10	19%	0.278	0.003		6%	0.006	0.987	
	ETV5/DGKG	rs1516728	2	Major	T	A	0.11	23%	-0.060	0.489		52%	-0.098	0.565	
4	GNPDA2	rs12641981	2	Minor	T	C	0.18	43%	0.088	0.238		23%	0.103	0.602	
	SLC39A8	rs13114738	1	Minor	T	C	0.13	8%	0.506	4.15E-04		1%	-1.583	0.008	X
5	POC5 FLJ35779	rs10057967	1	Major	C	T	0.10	37%	-0.227	0.003		49%	0.128	0.435	X
	ZNF608	rs6864049	1	Minor	G	A	0.07	46%	-0.189	0.012	X	19%	-0.463	0.033	X
6	TFAP2B	rs734597	3	Minor	A	G	0.13	17%	0.382	1.21E-04		9%	0.030	0.920	
9	LING02 LRRN6C	rs1412235	1	Minor	C	G	0.11	31%	0.003	0.970		16%	0.365	0.111	
	LMX1B	rs867559	2	Minor	G	A	0.24	20%	0.088	0.339		32%	0.025	0.889	
11	RPL27A	rs2028882	1	Major	C	A	0.06	50%	-0.065	0.375		66%	0.116	0.515	X
	BDNF	rs10501087	2	Major	C	T	0.18	21%	-0.223	0.013		7%	-0.521	0.181	
	MTCH2	rs12419692	2	Minor	A	C	0.05	36%	0.146	0.059		9%	0.012	0.968	
12	BDCDIN3D, FAIM2	rs7138803	3	Minor	A	G	0.12	38%	0.164	0.033		17%	0.100	0.650	
13	MTIF3, GRF3A	rs1475219	1	Minor	C	T	0.09	21%	0.262	0.004		22%	-0.099	0.632	X
14	PRKD1	rs1440983	1	Minor	A	G	0.15	5%	0.266	0.129		23%	0.156	0.449	
	NRXN3	rs7144011	2	Minor	T	G	0.13	22%	0.165	0.064		24%	0.164	0.428	
15	MAP2K5	rs28670272	1	Major	G	A	0.13	23%	-0.212	0.014		41%	0.005	0.977	X
16	GPR5B	rs11639988	1	Major	G	A	0.17	15%	0.006	0.952	X	24%	-0.262	0.194	
	ATXN2L, TUFM, SH2B1	rs12443881	3	Minor	T	C	0.15	39%	-0.005	0.948	X	9%	-0.607	0.030	X
	FTO	rs9939609	11	Minor	A	T	0.38	41%	0.496	8.19E-11		48%	0.129	0.443	
18	MC4R	rs12970134	6	Minor	A	G	0.21	26%	0.209	0.012		13%	0.057	0.822	
19	KCTD15	rs11084753	3	Major	A	G	0.04	33%	-0.071	0.371		36%	0.197	0.270	X
	QPCTL	rs11083779	1	Major	C	T	0.07	4%	-0.227	0.196		11%	-0.267	0.294	
	ZC3H4 TMEM160	rs7250850	1	Major	G	C	0.09	29%	-0.174	0.032		80%	-0.343	0.124	

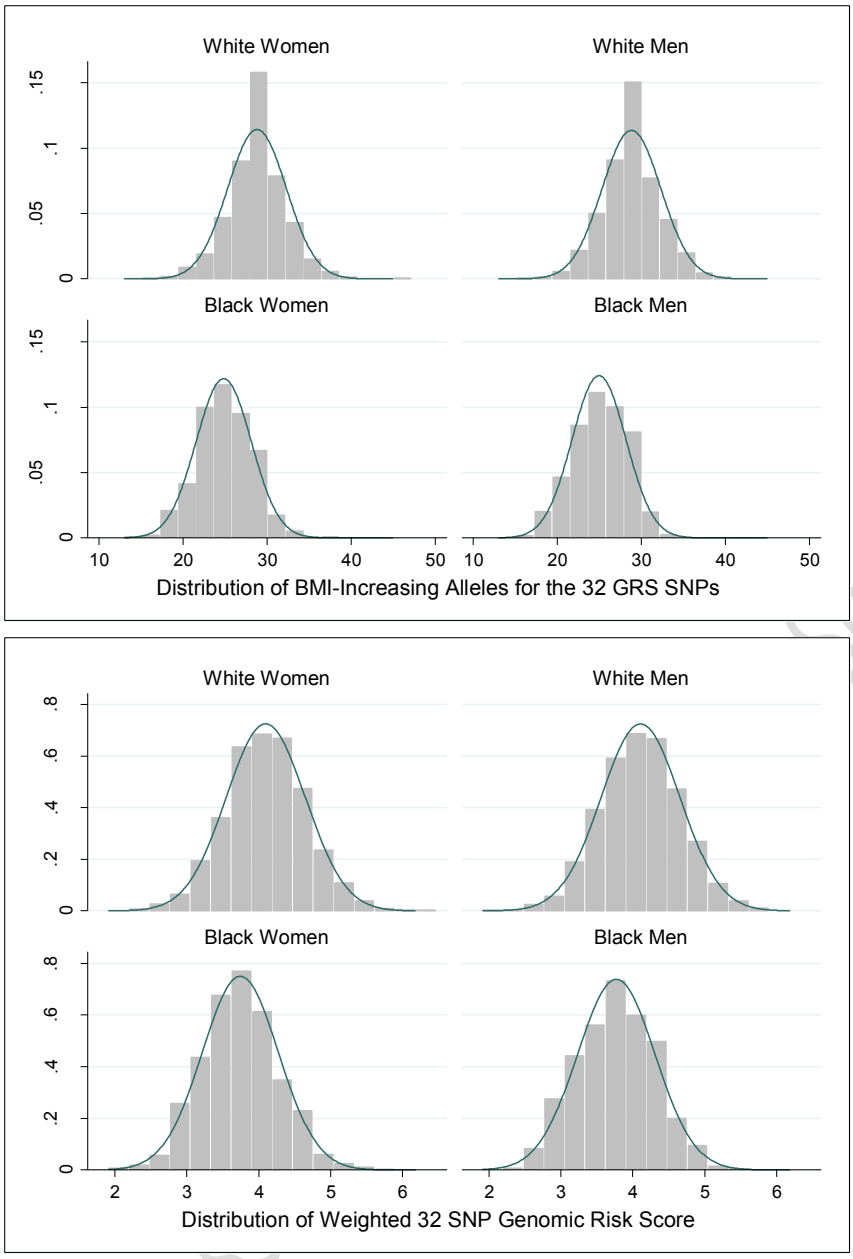
Supplementary Table 7 Footnote: GWAS replications include GWAS reporting any SNP in any LD block tagged by the SNP as obesity-associated at $p<1x10^{-5}$ in the discovery or combined discovery and replication samples. Test allele and other allele are reported from the positive strand. Effect-size weights were obtained from (Speliotes et al. 2010) for all SNPs with the exception of rs867559, for which the effect size weight was obtained from (Thorleifsson et al. 2009). Allele frequencies and per-allele effects are reported based on all participants in the analysis sample. Per-allele effects were estimated from linear regressions of BMI on SNP genotype (number of minor alleles), adjusted for demographics and geography. P-values are reported based on heteroskedasticity robust standard errors.

Supplementary Table 8. Educational Attainment of White and African American ARIC Participants. Educational attainment was ascertained via self-report at the first ARIC visit. Distributions of BMI-increasing alleles for the 32 obesity GRS SNPs were comparable across educational strata in African Americans and whites (p>0.10 for all comparisons).

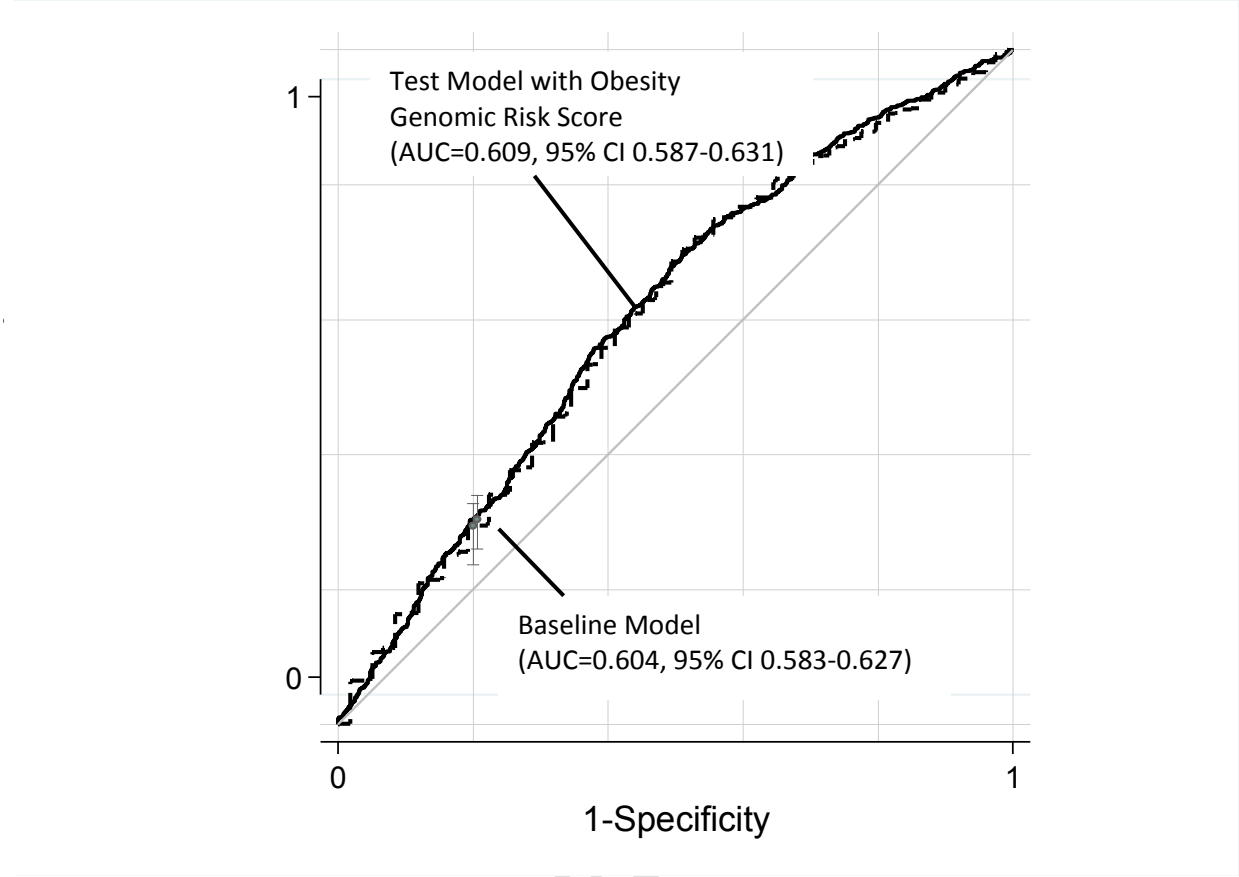
Highest Level of Schooling	Percent of Visit 1 Sample	
None/ Grade School	5%	19%
Some High School	11%	21%
High School Graduate	36%	22%
Vocational School	9%	7%
College	30%	18%
Graduate/ Professional School	9%	14%

Supplementary Table 9. Predictiveness of Model-Based Risk Scores With and Without The Obesity Genetic Risk Score. (m1-5) denote separate models used to estimate risk scores for BMI and obesity. Risk scores were predicted values from linear regression of BMI and predicted probabilities from probit regressions of obesity. The first model, m1, includes measures of age, sex, and ARIC Study Center where data were collected. The regression model was specified to include linear and quadratic terms for age and a product term modeling interaction between age and sex. The simple genetic risk assessment (SNPs in *FTO* and downstream of *MC4R*) is a component of the weighted obesity genomic risk score. Thus, model m3 contains all of the information in model m2 as well as information from the remaining 30 SNPs included in the GRS. The 5 categories of socioeconomic status were modeled as dichotomous variables and were allowed to vary by sex in their relationship with obesity and BMI. Values of R^2 were estimated using linear regression models adjusted for demographic and geographic information. Percentile-based confidence intervals were generated using the bootstrap method. AUCs and percentile-based confidence intervals were estimated from ROC curves constructed for predicted values generated using a probit regression model and were adjusted for the ARIC Study Center where data were collected using Pepe's method (Janes and Pepe 2009; Pepe, Longton, and Janes 2009). IDIs and test statistics were estimated only for comparisons of models m3 and m2 and models m5 and m4 using Pencina's Method (Pencina et al. 2008). IDIs for comparisons of models m2 and m3 with model m1 are identical to those reported for the respective obesity risk measures in Table 4 of the article.

		White ARIC Participants (n=8,286)			Black ARIC Participants (n=2,442)		
Model	Model Components	R^2 (95% CI)	AUC (95% CI)	IDI (p-value)	R^2 (95% CI)	AUC (95% CI)	IDI (p-value)
(m1)	Demographic & Geographic Information	3.20%	0.526		5.17%	0.604	
(m2)	m1 + Simple Genetic Risk Assessment	3.88%	0.550		5.35%	0.607	
(m3)	m1 + Weighted GRS	4.88%	0.574		5.52%	0.609	
	Change in predictiveness with addition of the weighted GRS	1.00% (0.006-0.014)	0.024 (0.012-0.036)	0.006 (7.81E-13)	0.17% (-0.001-0.005)	0.002 (-0.005-0.009)	0.001 (0.055)
(m4)	m1 + Socioeconomic Status	4.70%	0.550		7.70%	0.643	
(m5)	m4 + Weighted GRS	6.20%	0.586		7.92%	0.645	
	Change in predictiveness with addition of the weighted GRS	1.50% (0.010-0.020)	0.036 (0.023-0.050)	0.010 (5.46E-19)	0.22% (-0.001-0.006)	0.002 (-0.003-0.008)	0.002 (0.012)



Supplementary Figure 1. Distributions of BMI Increasing Alleles for the 32 GRS SNPs and the Weighted Obesity Genomic Risk Score Among White and African American ARIC Participants. Variance of the obesity genomic risk scores (GRS) was similar among women and men within ethnicity ($p>0.15$ for both samples), but was greater among whites as compared to African Americans ($p<0.001$) according to Brown and Forsythe's (Brown and Forsythe 1974) test for equality of variances.



Supplementary Figure 2. Receiver Operating Characteristic Curves for Obesity Among African American ARIC Participants (n=2,442). Baseline Model = gender, age (quadratic), gender x age interaction, ARIC study center; Test Model = baseline model + weighted obesity genomic risk score. ROC Curves were constructed using predicted values from probit regressions of obesity (BMI≥30) on the model terms. Delta AUC (AUC_{Test}-$AUC_{Baseline}$) = 0.005, 95% CI -0.005-0.015, p=0.30. Delta Partial AUC at 80% specificity=0, 95% CI -0.004-0.004, p=0.97. AUCs, partial AUCs, and delta AUCs were estimated using Pepe's method (Janes and Pepe 2009; Pepe, Longton, and Janes 2009).

References

Baltagi, B. H. 1980. On Seemingly Unrelated Regressions with Error-Components. *Econometrica* 48 (6):1547-1551.

Brown, M. B., and A. B. Forsythe. 1974. Robust Tests for Equality of Variances. *Journal of the American Statistical Association* 69 (346):364-367.

Cotsapas, C., E. K. Speliotes, I. J. Hatoum, D. M. Greenawalt, R. Dobrin, P. Y. Lum, C. Suver, E. Chudin, D. Kemp, M. Reitman, B. F. Voight, B. M. Neale, E. E. Schadt, J. N. Hirschhorn, L. M. Kaplan, and M. J. Daly. 2009. Common body mass index-associated variants confer risk of extreme obesity. *Human Molecular Genetics* 18 (18):3502-7.

Fox, C. S., N. Heard-Costa, L. A. Cupples, J. Dupuis, R. S. Vasan, and L. D. Atwood. 2007. Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. *BMC Medical Genetics* 8 Suppl 1:S18.

Frayling, T. M., N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. Perry, K. S. Elliott, H. Lango, N. W. Rayner, B. Shields, L. W. Harries, J. C. Barrett, S. Ellard, C. J. Groves, B. Knight, A. M. Patch, A. R. Ness, S. Ebrahim, D. A. Lawlor, S. M. Ring, Y. Ben-Shlomo, M. R. Jarvelin, U. Sovio, A. J. Bennett, D. Melzer, L. Ferrucci, R. J. Loos, I. Barroso, N. J. Wareham, F. Karpe, K. R. Owen, L. R. Cardon, M. Walker, G. A. Hitman, C. N. Palmer, A. S. Doney, A. D. Morris, G. D. Smith, A. T. Hattersley, and M. I. McCarthy. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316 (5826):889-94.

GENEVA ARIC Project. 2009. Quality Control Report for the ARIC GWAS database. Bethesda, MD: The National Institutes of Health Database of Genotypes and Phenotypes (dbGaP).

Heard-Costa, N. L., M. C. Zillikens, K. L. Monda, A. Johansson, T. B. Harris, M. Fu, T. Haritunians, M. F. Feitosa, T. Aspelund, G. Eiriksdottir, M. Garcia, L. J. Launer, A. V. Smith, B. D. Mitchell, P. F. McArdle, A. R. Shuldiner, S. J. Bielinski, E. Boerwinkle, F. Brancati, E. W. Demerath, J. S. Pankow, A. M. Arnold, Y. D. Chen, N. L. Glazer, B. McKnight, B. M. Psaty, J. I. Rotter, N. Amin, H. Campbell, U. Gyllensten, C. Pattaro, P. P. Pramstaller, I. Rudan, M. Struchalin, V. Vitart, X. Gao, A. Kraja, M. A. Province, Q. Zhang, L. D. Atwood, J. Dupuis, J. N. Hirschhorn, C. E. Jaquish, C. J. O'Donnell, R. S. Vasan, C. C. White, Y. S. Aulchenko, K. Estrada, A. Hofman, F. Rivadeneira, A. G. Uitterlinden, J. C. Witteman, B. A. Oostra, R. C. Kaplan, V. Gudnason, J. R. O'Connell, I. B. Borecki, C. M. van Duijn, L. A. Cupples, C. S. Fox, and K. E. North. 2009. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genetics* 5 (6):e1000539.

Herbert, A., N. P. Gerry, M. B. McQueen, I. M. Heid, A. Pfeufer, T. Illig, H. E. Wichmann, T. Meitinger, D. Hunter, F. B. Hu, G. Colditz, A. Hinney, J. Hebebrand, K. Koberwitz, X. Zhu, R. Cooper, K. Ardlie, H. Lyon, J. N. Hirschhorn, N. M. Laird, M. E. Lenburg, C. Lange, and M. F. Christman. 2006. A common genetic variant is associated with adult and childhood obesity. *Science* 312 (5771):279-83.

Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 106 (23):9362-7.

Hinney, A., T. T. Nguyen, A. Scherag, S. Friedel, G. Bronner, T. D. Muller, H. Grallert, T. Illig, H. E. Wichmann, W. Rief, H. Schafer, and J. Hebebrand. 2007. Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PloS One* 2 (12):e1361.

Janes, H., and M. S. Pepe. 2009. Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika* 96 (2):371-382.

Johansson, A., F. Marroni, C. Hayward, C. S. Franklin, A. V. Kirichenko, I. Jonasson, A. A. Hicks, V. Vitart, A. Isaacs, T. Axenovich, S. Campbell, J. Floyd, N. Hastie, S. Knott, G. Lauc, I. Pichler, K. Rotim, S. H. Wild, I. V. Zorkoltseva, J. F. Wilson, I. Rudan, H. Campbell, C. Pattaro, P. Pramstaller, B. A. Oostra, A. F. Wright, C. M. van Duijn, Y. S. Aulchenko, U. Gyllensten, and Eurospan Consortium. 2010. Linkage and genome-wide association analysis of obesity-related phenotypes: Association of weight with the MGAT1 gene. *Obesity* 18 (4):803-808.

Li, S., J. H. Zhao, J. Luan, R. N. Luben, S. A. Rodwell, K. T. Khaw, K. K. Ong, N. J. Wareham, and R. J. Loos. 2010. Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *American Journal of Clinical Nutrition* 91 (1):184-90.

Lindgren, C. M., I. M. Heid, J. C. Randall, C. Lamina, V. Steinthorsdottir, L. Qi, E. K. Speliotes, G. Thorleifsson, C. J. Willer, B. M. Herrera, A. U. Jackson, N. Lim, P. Scheet, N. Soranzo, N. Amin, Y. S. Aulchenko, J. C. Chambers, A. Drong, J. A. Luan, H. N. Lyon, F. Rivadeneira, S. Sanna, N. J. Timpson, M. C. Zillikens, J. H. Zhao, P. Almgren, S. Bandinelli, A. J. Bennett, R. N. Bergman, L. L. Bonnycastle, S. J. Bumpstead, S. J. Chanock, L. Cherkas, P. Chines, L. Coin, C. Cooper, G. Crawford, A. Doering, A. Dominiczak, A. S. F. Doney, S. Ebrahim, P. Elliott, M. R. Erdos, K. Estrada, L. Ferrucci, G. Fischer, N. G. Forouhi, C. Gieger, H. Grallert, C. J. Groves, S. Grundy, C. Guiducci, D. Hadley, A. Hamsten, A. S. Havulinna, A. Hofman, R. Holle, J. W. Holloway, T. Illig, B. Isomaa, L. C. Jacobs, K. Jameson, P. Jousilahti, F. Karpe, J. Kuusisto, J. Laitinen, G. M. Lathrop, D. A. Lawlor, M. Mangino, W. L. McArdle, T. Meitinger, M. A. Morken, A. P. Morris, P. Munroe, N. Narisu, A. Nordstrom, P. Nordstrom, B. A. Oostra, C. N. A. Palmer, F. Payne, J. F. Peden, I. Prokopenko, F. Renstrom, A. Ruokonen, V. Salomaa, M. S. Sandhu, L. J. Scott, A. Scuteri, K. Silander, K. J. Song, X. Yuan, H. M. Stringham, A. J. Swift, T. Tuomi, M. Uda, P. Vollenweider, G. Waeber, C. Wallace, G. B. Walters, M. N. Weedon, J. C. M. Witteman, C. L. Zhang, W. H. Zhang, M. J. Caulfield, F. S. Collins, G. D. Smith, I. N. M. Day, P. W. Franks, A. T. Hattersley, F. B. Hu, M. R. Jarvelin, A. Kong, J. S. Kooner, M. Laakso, E. Lakatta, V. Mooser, A. D. Morris, L. Peltonen, N. J. Samani, T. D. Spector, D. P. Strachan, T. Tanaka, J. Tuomilehto, A. G. Uitterlinden, C. M. van Duijn, N. J. Wareham, H. Watkins, D. M. Waterworth, M. Boehnke, P. Deloukas, L. Groop, D. J. Hunter, U. Thorsteinsdottir, D. Schlessinger, H. E. Wichmann, T. M. Frayling, G. R. Abecasis, J. N. Hirschhorn, R. J. F. Loos, K. Stefansson, K. L. Mohlke, I. S. Barroso, M. I. McCarthy, Procardis Consortia Giant Wellcome Trust Case Control Consor, and Consortium. 2009. Genome-Wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genetics* 5 (6):e1000508.

Liu, J. Z., S. E. Medland, M. J. Wright, A. K. Henders, A. C. Heath, P. A. Madden, A. Duncan, G. W. Montgomery, N. G. Martin, and A. F. McRae. 2010. Genome-wide association study of height and body mass index in Australian twin families. *Twin Research and Human Genetics* 13 (2):179-93.

Liu, Y. J., X. G. Liu, L. Wang, C. Dina, H. Yan, J. F. Liu, S. Levy, C. J. Papasian, B. M. Drees, J. J. Hamilton, D. Meyre, J. Delplanque, Y. F. Pei, L. Zhang, R. R. Recker, P. Froguel, and H. W. Deng. 2008. Genome-wide association scans identified CTNNBL1 as a novel gene for obesity. *Human Molecular Genetics* 17 (12):1803-13.

Loos, R. J., C. M. Lindgren, S. Li, E. Wheeler, J. H. Zhao, I. Prokopenko, M. Inouye, R. M. Freathy, A. P. Attwood, J. S. Beckmann, S. I. Berndt, K. B. Jacobs, S. J. Chanock, R. B. Hayes, S. Bergmann, A. J. Bennett, S. A. Bingham, M. Bochud, M. Brown, S. Cauchi, J. M. Connell, C. Cooper, G. D. Smith, I. Day, C. Dina, S. De, E. T. Dermitzakis, A. S. Doney, K. S. Elliott, P. Elliott, D. M. Evans, I. Sadaf Farooqi, P. Froguel, J. Ghori, C. J. Groves, R. Gwilliam, D. Hadley, A. S. Hall, A. T. Hattersley, J.

Hebebrand, I. M. Heid, C. Lamina, C. Gieger, T. Illig, T. Meitinger, H. E. Wichmann, B. Herrera, A. Hinney, S. E. Hunt, M. R. Jarvelin, T. Johnson, J. D. Jolley, F. Karpe, A. Keniry, K. T. Khaw, R. N. Luben, M. Mangino, J. Marchini, W. L. McArdle, R. McGinnis, D. Meyre, P. B. Munroe, A. D. Morris, A. R. Ness, M. J. Neville, A. C. Nica, K. K. Ong, S. O'Rahilly, K. R. Owen, C. N. Palmer, K. Papadakis, S. Potter, A. Pouta, L. Qi, J. C. Randall, N. W. Rayner, S. M. Ring, M. S. Sandhu, A. Scherag, M. A. Sims, K. Song, N. Soranzo, E. K. Speliotes, H. E. Syddall, S. A. Teichmann, N. J. Timpson, J. H. Tobias, M. Uda, C. I. Vogel, C. Wallace, D. M. Waterworth, M. N. Weedon, C. J. Willer, Wraight, X. Yuan, E. Zeggini, J. N. Hirschhorn, D. P. Strachan, W. H. Ouwehand, M. J. Caulfield, N. J. Samani, T. M. Frayling, P. Vollenweider, G. Waeber, V. Mooser, P. Deloukas, M. I. McCarthy, N. J. Wareham, I. Barroso, P. Kraft, S. E. Hankinson, D. J. Hunter, F. B. Hu, H. N. Lyon, B. F. Voight, M. Ridderstrale, L. Groop, P. Scheet, S. Sanna, G. R. Abecasis, G. Albai, R. Nagaraja, D. Schlessinger, A. U. Jackson, J. Tuomilehto, F. S. Collins, M. Boehnke, and K. L. Mohlke. 2008. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genetics* 40 (6):768-75.

Meyre, D., J. Delplanque, J. C. Chevre, C. Lecoeur, S. Lobbens, S. Gallina, E. Durand, V. Vatin, F. Degraeve, C. Proenca, S. Gaget, A. Korner, P. Kovacs, W. Kiess, J. Tichet, M. Marre, A. L. Hartikainen, F. Horber, N. Potoczna, S. Hercberg, C. Levy-Marchal, F. Pattou, B. Heude, M. Tauber, M. I. McCarthy, A. I. Blakemore, A. Montpetit, C. Polychronakos, J. Weill, L. J. Coin, J. Asher, P. Elliott, M. R. Jarvelin, S. Visvikis-Siest, B. Balkau, R. Sladek, D. Balding, A. Walley, C. Dina, and P. Froguel. 2009. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nature Genetics* 41 (2):157-9.

Pencina, M. J., R. B. D'Agostino, Sr., R. B. D'Agostino, Jr., and R. S. Vasan. 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine* 27 (2):157-72; discussion 207-12.

Pepe, M., G. Longton, and H. Janes. 2009. Estimation and Comparison of Receiver Operating Characteristic Curves. *Stata Journal* 9 (1):1.

Peterson, R. E., H. H. Maes, P. Holmans, A. R. Sanders, D. F. Levinson, J. Shi, K. S. Kendler, P. V. Gejman, and B. T. Webb. 2011. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Human Genetics* 129 (2):221-30.

Royston, P., and G. Ambler. 1999. Multivariable fractional polynomials: Update. In *Stata Technical Bulletin*. College Station, Tx: Stata.

Royston, P., and W. Sauerbrei. 2003. Stability of multivariable fractional polynomial models with selection of variables and transformations: a bootstrap investigation. *Statistics in Medicine* 22 (4):639-59.

Scherag, A., C. Dina, A. Hinney, V. Vatin, S. Scherag, C. I. Vogel, T. D. Muller, H. Grallert, H. E. Wichmann, B. Balkau, B. Heude, M. R. Jarvelin, A. L. Hartikainen, C. Levy-Marchal, J. Weill, J. Delplanque, A. Korner, W. Kiess, P. Kovacs, N. W. Rayner, I. Prokopenko, M. I. McCarthy, H. Schafer, I. Jarick, H. Boeing, E. Fisher, T. Reinehr, J. Heinrich, P. Rzehak, D. Berdel, M. Borte, H. Biebermann, H. Krude, D. Rosskopf, C. Rimmbach, W. Rief, T. Fromme, M. Klingenspor, A. Schurmann, N. Schulz, M. M. Nothen, T. W. Muhleisen, R. Erbel, K. H. Jockel, S. Moebus, T. Boes, T. Illig, P. Froguel, J. Hebebrand, and D. Meyre. 2010. Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. *PLoS Genetics* 6 (4):e1000916.

Scuteri, A., S. Sanna, W. M. Chen, M. Uda, G. Albai, J. Strait, S. Najjar, R. Nagaraja, M. Orru, G. Usala, M. Dei, S. Lai, A. Maschio, F. Busonero, A. Mulas, G. B. Ehret, A. A. Fink, A. B. Weder, R. S. Cooper, P. Galan, A. Chakravarti, D. Schlessinger, A. Cao, E. Lakatta, and G. R. Abecasis. 2007. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genetics* 3 (7):1200-1210.

Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, C. M. Lindgren, J. Luan, R. Magi, J. C. Randall, S. Vedantam, T. W. Winkler, L. Qi, T. Workalemahu, I. M. Heid, V. Steinthorsdottir, H. M. Stringham, M. N. Weedon, E. Wheeler, A. R. Wood, T. Ferreira, R. J. Weyant, A. V. Segre, K. Estrada, L. Liang, J. Nemesh, J. H. Park, S. Gustafsson, T. O. Kilpelainen, J. Yang, N. Bouatia-Naji, T. Esko, M. F. Feitosa, Z. Kutalik, M. Mangino, S. Raychaudhuri, A. Scherag, A. V. Smith, R. Welch, J. H. Zhao, K. K. Aben, D. M. Absher, N. Amin, A. L. Dixon, E. Fisher, N. L. Glazer, M. E. Goddard, N. L. Heard-Costa, V. Hoesel, J. J. Hottenga, A. Johansson, T. Johnson, S. Ketkar, C. Lamina, S. Li, M. F. Moffatt, R. H. Myers, N. Narisu, J. R. Perry, M. J. Peters, M. Preuss, S. Ripatti, F. Rivadeneira, C. Sandholt, L. J. Scott, N. J. Timpson, J. P. Tyrer, S. van Wingerden, R. M. Watanabe, C. C. White, F. Wiklund, C. Barlassina, D. I. Chasman, M. N. Cooper, J. O. Jansson, R. W. Lawrence, N. Pellikka, I. Prokopenko, J. Shi, E. Thiering, H. Alavere, M. T. Alibrandi, P. Almgren, A. M. Arnold, T. Aspelund, L. D. Atwood, B. Balkau, A. J. Balmforth, A. J. Bennett, Y. Ben-Shlomo, R. N. Bergman, S. Bergmann, H. Biebermann, A. I. Blakemore, T. Boes, L. L. Bonnycastle, S. R. Bornstein, M. J. Brown, T. A. Buchanan, F. Busonero, H. Campbell, F. P. Cappuccio, C. Cavalcanti-Proenca, Y. D. Chen, C. M. Chen, P. S. Chines, R. Clarke, L. Coin, J. Connell, I. N. Day, M. Heijer, J. Duan, S. Ebrahim, P. Elliott, R. Elosua, G. Eiriksdottir, M. R. Erdos, J. G. Eriksson, M. F. Facheris, S. B. Felix, P. Fischer-Posovszky, A. R. Folsom, N. Friedrich, N. B. Freimer, M. Fu, S. Gaget, P. V. Gejman, E. J. Geus, C. Gieger, A. P. Gjesing, A. Goel, P. Goyette, H. Grallert, J. Grassler, D. M. Greenawalt, C. J. Groves, V. Gudnason, C. Guiducci, A. L. Hartikainen, N. Hassanali, A. S. Hall, A. S. Havulinna, C. Hayward, A. C. Heath, C. Hengstenberg, A. A. Hicks, A. Hinney, A. Hofman, G. Homuth, J. Hui, W. Igl, C. Iribarren, B. Isomaa, K. B. Jacobs, I. Jarick, E. Jewell, U. John, T. Jorgensen, P. Jousilahti, A. Jula, M. Kaakinen, E. Kajantie, L. M. Kaplan, S. Kathiresan, J. Kettunen, L. Kinnunen, J. W. Knowles, I. Kolcic, I. R. Konig, S. Koskinen, P. Kovacs, J. Kuusisto, P. Kraft, K. Kvaloy, J. Laitinen, O. Lantieri, C. Lanzani, L. J. Launer, C. Lecoeur, T. Lehtimaki, G. Lettre, J. Liu, M. L. Lokki, M. Lorentzon, R. N. Luben, B. Ludwig, P. Manunta, D. Marek, M. Marre, N. G. Martin, W. L. McArdle, A. McCarthy, B. McKnight, T. Meitinger, O. Melander, D. Meyre, K. Midthjell, G. W. Montgomery, M. A. Morken, A. P. Morris, R. Mulic, J. S. Ngwa, M. Nelis, M. J. Neville, D. R. Nyholt, C. J. O'Donnell, S. O'Rahilly, K. K. Ong, B. Oostra, G. Pare, A. N. Parker, M. Perola, I. Pichler, K. H. Pietilainen, C. G. Platou, O. Polasek, A. Pouta, S. Rafelt, O. Raitakari, N. W. Rayner, M. Ridderstrale, W. Rief, A. Ruokonen, N. R. Robertson, P. Rzehak, V. Salomaa, A. R. Sanders, M. S. Sandhu, S. Sanna, J. Saramies, M. J. Savolainen, S. Scherag, S. Schipf, S. Schreiber, H. Schunkert, K. Silander, J. Sinisalo, D. S. Siscovick, J. H. Smit, N. Soranzo, U. Sovio, J. Stephens, I. Surakka, A. J. Swift, M. L. Tammesoo, J. C. Tardif, M. Teder-Laving, T. M. Teslovich, J. R. Thompson, B. Thomson, A. Tonjes, T. Tuomi, J. B. van Meurs, G. J. van Ommen, V. Vatin, J. Viikari, S. Visvikis-Siest, V. Vitart, C. I. Vogel, B. F. Voight, L. L. Waite, H. Wallaschofski, G. B. Walters, E. Widen, S. Wiegand, S. H. Wild, G. Willemsen, D. R. Witte, J. C. Witteman, J. Xu, Q. Zhang, L. Zgaga, A. Ziegler, P. Zitting, J. P. Beilby, I. S. Farooqi, J. Hebebrand, H. V. Huikuri, A. L. James, M. Kahonen, D. F. Levinson, F. Macciardi, M. S. Nieminen, C. Ohlsson, L. J. Palmer, P. M. Ridker, M. Stumvoll, J. S. Beckmann, H. Boeing, E. Boerwinkle, D. I. Boomsma, M. J. Caulfield, S. J. Chanock, F. S. Collins, L. A. Cupples, G. D. Smith, J. Erdmann, P. Froguel, H. Gronberg, U. Gyllensten, P. Hall, T. Hansen, T. B. Harris, A. T. Hattersley, R. B. Hayes, J. Heinrich, F. B. Hu, K. Hveem, T. Illig, M. R. Jarvelin, J. Kaprio, F. Karpe, K. T. Khaw, L. A. Kiemeney, H. Krude, M. Laakso, D. A. Lawlor, A. Metspalu, P. B. Munroe, W. H. Ouwehand, O. Pedersen, B. W. Penninx, A. Peters, P. P. Pramstaller, T. Quertermous, T. Reinehr, A. Rissanen, I. Rudan, N. J. Samani, P. E. Schwarz, A. R. Shuldiner, T. D. Spector, J. Tuomilehto, M. Uda, A. Uitterlinden, T. T. Valle, M. Wabitsch, G. Waeber, N. J. Wareham, H. Watkins, J. F. Wilson, A. F. Wright, M. C. Zillikens, N. Chatterjee, S. A. McCarroll, S. Purcell, E. E. Schadt, P. M. Visscher, T. L. Assimes, I. B. Borecki, P. Deloukas, C. S. Fox, L. C. Groop, T. Haritunians, D. J. Hunter, R. C. Kaplan, K. L.

Mohlke, J. R. O'Connell, L. Peltonen, D. Schlessinger, D. P. Strachan, C. M. van Duijn, H. E. Wichmann, T. M. Frayling, U. Thorsteinsdottir, G. R. Abecasis, I. Barroso, M. Boehnke, K. Stefansson, K. E. North, M. I. McCarthy, J. N. Hirschhorn, E. Ingelsson, and R. J. Loos. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* 42 (11):937-48.

Thorleifsson, G., G. B. Walters, D. F. Gudbjartsson, V. Steinthorsdottir, P. Sulem, A. Helgadottir, U. Styrkarsdottir, S. Gretarsdottir, S. Thorlacius, I. Jonsdottir, T. Jonsdottir, E. J. Olafsdottir, G. H. Olafsdottir, T. Jonsson, F. Jonsson, K. Borch-Johnsen, T. Hansen, G. Andersen, T. Jorgensen, T. Lauritzen, K. K. Aben, A. L. Verbeek, N. Roeleveld, E. Kampman, L. R. Yanek, L. C. Becker, L. Tryggvadottir, T. Rafnar, D. M. Becker, J. Gulcher, L. A. Kiemeney, O. Pedersen, A. Kong, U. Thorsteinsdottir, and K. Stefansson. 2009. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nature Genetics* 41 (1):18-24.

Verzilli, C. J., N. Stallard, and J. C. Whittaker. 2005. Bayesian modelling of multivariate quantitative traits using seemingly unrelated regressions. *Genetic Epidemiology* 28 (4):313-25.

Willer, C. J., E. K. Speliotes, R. J. Loos, S. Li, C. M. Lindgren, I. M. Heid, S. I. Berndt, A. L. Elliott, A. U. Jackson, C. Lamina, G. Lettre, N. Lim, H. N. Lyon, S. A. McCarroll, K. Papadakis, L. Qi, J. C. Randall, R. M. Roccasecca, S. Sanna, P. Scheet, M. N. Weedon, E. Wheeler, J. H. Zhao, L. C. Jacobs, I. Prokopenko, N. Soranzo, T. Tanaka, N. J. Timpson, P. Almgren, A. Bennett, R. N. Bergman, S. A. Bingham, L. L. Bonnycastle, M. Brown, N. P. Burtt, P. Chines, L. Coin, F. S. Collins, J. M. Connell, C. Cooper, G. D. Smith, E. M. Dennison, P. Deodhar, P. Elliott, M. R. Erdos, K. Estrada, D. M. Evans, L. Gianniny, C. Gieger, C. J. Gillson, C. Guiducci, R. Hackett, D. Hadley, A. S. Hall, A. S. Havulinna, J. Hebebrand, A. Hofman, B. Isomaa, K. B. Jacobs, T. Johnson, P. Jousilahti, Z. Jovanovic, K. T. Khaw, P. Kraft, M. Kuokkanen, J. Kuusisto, J. Laitinen, E. G. Lakatta, J. Luan, R. N. Luben, M. Mangino, W. L. McArdle, T. Meitinger, A. Mulas, P. B. Munroe, N. Narisu, A. R. Ness, K. Northstone, S. O'Rahilly, C. Purmann, M. G. Rees, M. Ridderstrale, S. M. Ring, F. Rivadeneira, A. Ruokonen, M. S. Sandhu, J. Saramies, L. J. Scott, A. Scuteri, K. Silander, M. A. Sims, K. Song, J. Stephens, S. Stevens, H. M. Stringham, Y. C. Tung, T. T. Valle, C. M. Van Duijn, K. S. Vimaleswaran, P. Vollenweider, G. Waeber, C. Wallace, R. M. Watanabe, D. M. Waterworth, N. Watkins, J. C. Witteman, E. Zeggini, G. Zhai, M. C. Zillikens, D. Altshuler, M. J. Caulfield, S. J. Chanock, I. S. Farooqi, L. Ferrucci, J. M. Guralnik, A. T. Hattersley, F. B. Hu, M. R. Jarvelin, M. Laakso, V. Mooser, K. K. Ong, W. H. Ouwehand, V. Salomaa, N. J. Samani, T. D. Spector, T. Tuomi, J. Tuomilehto, M. Uda, A. G. Uitterlinden, N. J. Wareham, P. Deloukas, T. M. Frayling, L. C. Groop, R. B. Hayes, D. J. Hunter, K. L. Mohlke, L. Peltonen, D. Schlessinger, D. P. Strachan, H. E. Wichmann, M. I. McCarthy, M. Boehnke, I. Barroso, G. R. Abecasis, and J. N. Hirschhorn. 2009. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics* 41 (1):25-34.