

## Genomewide Linkage Analysis of Body Mass Index across 28 Years of the Framingham Heart Study

Larry D. Atwood,<sup>1,3</sup> Nancy L. Heard-Costa,<sup>1</sup> L. Adrienne Cupples,<sup>3</sup> Cashell E. Jaquish,<sup>5</sup> Peter W. F. Wilson,<sup>2</sup> and Ralph B. D'Agostino<sup>4</sup>

Departments of <sup>1</sup>Neurology and <sup>2</sup>Medicine, School of Medicine, Department of <sup>3</sup>Biostatistics, School of Public Health, and <sup>4</sup>Department of Mathematics and Statistics, Boston University, Boston; and <sup>5</sup>National Heart, Lung, and Blood Institute, Bethesda

We performed a genomewide linkage analysis of six separate measurements of body mass index (BMI) taken over a span of 28 years, from 1971 to 1998, in the Framingham Heart Study. Variance-components linkage analysis was performed on 330 families, using 401 polymorphic markers. The number of individuals with data at each exam ranged from 1,930, in 1971, to 1,401, in 1998. Sex, age, and age squared were included as covariates in the model. There was substantial evidence for linkage on chromosome 6q23-25, in the area of D6S1009, GATA184A08, D6S2436, and D6S305. The six measurements had maximum LOD scores of 4.64, 2.29, 2.41, 1.40, 0.99, and 3.08, respectively, all in the chromosome 6q23-25 region. There was also evidence for linkage of multiple measures on chromosome 11q14 in the area of D11S1998, D11S4464, and D11S912. The six measurements had maximum LOD scores of 0.61, 3.27, 1.30, 0.68, 1.30, and 2.29, respectively, all in the chromosome 11q14 region. Both of these regions have been reported in previous studies. Evidence in the same regions from multiple measurements does not constitute replication; however, it does indicate that linkage studies of BMI are robust with respect to measurement error. It is unclear whether the variation in LOD scores in these regions is due to age effects, varying sample size, or other confounding factors.

### Introduction

BMI is one of the most heavily studied measures of obesity (MIM #601665). The scientific literature on the genetics of obesity is extensive and has been superbly reviewed by Rankinen et al. (2002), in the latest of a series of reviews on the genetics of obesity. They report that the total number of genes, markers, or chromosomal regions that have been associated or linked with obesity in humans now exceeds 250. In particular, there are 33 QTLs in humans, with varying levels of support, based on published genome scans. Since that review went to press, six genome scans of BMI have been reported; four found significant linkage and one found suggestive linkage. Hunt et al. (2001) found significant linkage of BMI to chromosome 20, in a large set of Utah families ascertained for obesity. Feitosa et al. (2002), in a large population-based sample of families, found strong evidence for linkage of BMI to chromosomes 7q32.3 and 13q14. Wu et al. (2002) found significant linkage to chromosome 3q27, in a large sample of hypertensive families. Zhu et al. (2002) found suggestive linkage for

BMI on chromosome 5p15. All five of these regions on chromosomes 3, 5, 7, 13, and 20 have been previously reported and are reviewed by Rankinen et al. (2002). Deng et al. (2002) found significant linkage to chromosome 2q14 in a sample of relative pairs selected for low bone mineral density. Of these five recent studies, this result is the only new linkage region for BMI. The sixth recent genomewide scan was a search in the Pima Indians for evidence of imprinting (Lindsay et al. 2001), for which little evidence was found.

Two methodological issues, which are rarely addressed, are those of robustness to measurement error and genetic change over time. All quantitative traits will have some error that is due to the imperfection of the measurement procedure and that contributes to phenotypic variation. A direct test of the contribution of measurement error is to take multiple independent measurements on the same group of individuals and repeat the genetic analysis on each of the multiple measures. Furthermore, the relative contribution of a gene to a specific trait could change over the lifetime of an individual—that is, the genetic effect could be age-specific. A direct test of the genetic change with age is to take multiple independent measurements on the same group of individuals as they age and to repeat the genetic analysis on each of the multiple measures. Unfortunately, most genetic studies are cross-sectional and do not have multiple measurements across time. Thus, the effects of

Received June 21, 2002; accepted for publication July 29, 2002; electronically published September 27, 2002.

Address for correspondence and reprints: Dr. Larry D. Atwood, Department of Neurology, Boston University School of Medicine, 715 Albany Street, B-609, Boston, MA 02118. E-mail: lda@bu.edu

© 2002 by The American Society of Human Genetics. All rights reserved. 0002-9297/2002/7105-0005\$15.00

measurement error and aging cannot be evaluated in cross-sectional studies.

An advantage of the Framingham Heart Study (FHS) is its repeated-measure prospective design; it is not limited to the cross-sectional design of most genetic studies. The FHS has observations on families taken over the course of 28 years, beginning when the offspring of the original cohort were recruited, in 1971–1975. Between 1971 and 1998, most participants from these families were measured at least six times. BMI might be an informative trait for a study of the change with age, since the basis of BMI (i.e., height and weight) can be measured with minimal error. However, measurement error still exists. If no common regions across time are observed or no obvious age-related patterns emerge, then it may be that the genetic effects are too small or the linkage methodology is sensitive to even the minimal measurement error inherent in BMI, or it may be that the genetic effects of age are highly complex. Another potentially confounding factor is the varying sample size, which is primarily due to death of the participants.

In the present study, we observed common linkage regions across time for two regions on chromosomes 6 and 11. The LOD scores in these two regions range from minimally significant to highly significant. It is unclear whether this variation in LOD scores is due to genetic change with age or to measurement error.

**Subjects and Methods**

*Design*

The FHS has been described in detail elsewhere (Cupples et al. 1988). In brief, the FHS divides subjects into two recruitment groups. The first group is the original cohort of subjects, referred to as the “cohort” here. This cohort consisted of the adult members between the ages of 28 and 62 years in about two-thirds of the households in the town of Framingham, Massachusetts, in 1948. Since then, the cohort has been examined every two years, for a total of 26 exams. The second group primarily comprises the children of the original cohort and their spouses and is referred to as the “offspring” here. The offspring were first examined in 1971–1974. They

were next examined in 1979–1982 and every four years thereafter up to 1998, for a total of six exams. In the mid-1990s, DNA was extracted and shipped to the Mammalian Genotyping Service (MGS). The MGS produced genotypes for 1,702 individuals in 330 families, for 401 polymorphic markers (marker set 8A; average heterozygosity 0.77) with an average intermarker spacing of 8.59 cM (SD = 3.97). The genotypes were checked for Mendelian consistency by the PEDSYS program (Dyke 1996).

At each exam, each subject undergoes an extensive data-gathering protocol. For the present study, we use only five variables: sex, age, height, weight, and smoking status. BMI was derived as weight (in kilograms) divided by the square of height (in meters).

We constructed six data sets corresponding to the six offspring exams. Each data set consisted of 330 pedigrees comprising 546 sibships ranging in size from 0 to 7. For each individual in each data set, sex, age, and BMI were selected as follows: in the first data set, the variables were taken from offspring exam 1 and cohort exam 10 (1966–1970), depending on whether the individual was a member of the cohort or the offspring. Offspring exam 1 and cohort exam 10 come closest to overlapping in time (height was not measured in closer cohort exams) and are thus the natural choice to achieve a cross-sectional data set of these families. For data sets 2–6, we combined offspring 2 and cohort 16 (1979–1982), offspring 3 and cohort 18 (1983–1987), offspring 4 and cohort 20 (1987–1990), offspring 5 and cohort 22 (1991–1994), and offspring 6 and cohort 24 (1995–1998), respectively. It should be emphasized that the data sets are numbered to indicate their temporal relationship; the individuals are youngest in data set 1 and oldest in data set 6. All subjects gave informed consent, and the institutional review board of the Boston University School of Medicine has approved all data-gathering protocols.

*Statistical Analysis*

Linkage analysis was performed using a variance-components approach (Amos 1994; Almasy and Blangero 1998), as implemented in Genehunter (Kruglyak et al.

**Table 1**  
**Descriptive Statistics for All Six Data sets**

DATA SET	N	% FEMALE	AGE IN YEARS (MEAN ± SD)	BMI			
				Mean ± SD	<i>h</i> <sup>2</sup>	Skewness	Kurtosis
1 (1971–1975)	1,930	51.0	40.4 ± 15.1	25.5 ± 4.1	.37	.6	.6
2 (1979–1982)	1,764	52.8	50.1 ± 15.7	26.0 ± 4.3	.38	.8	1.1
3 (1983–1987)	1,678	52.4	52.6 ± 15.1	26.4 ± 4.6	.38	.9	1.6
4 (1987–1990)	1,679	52.1	54.7 ± 14.8	26.9 ± 4.8	.45	1.0	1.7
5 (1991–1994)	1,546	52.8	57.5 ± 14.3	27.5 ± 5.0	.47	1.0	1.6
6 (1994–1998)	1,401	52.7	60.0 ± 13.3	28.0 ± 5.2	.52	1.0	1.7

**Table 2****All Maximum LOD Scores >1.0 for All Data Sets**

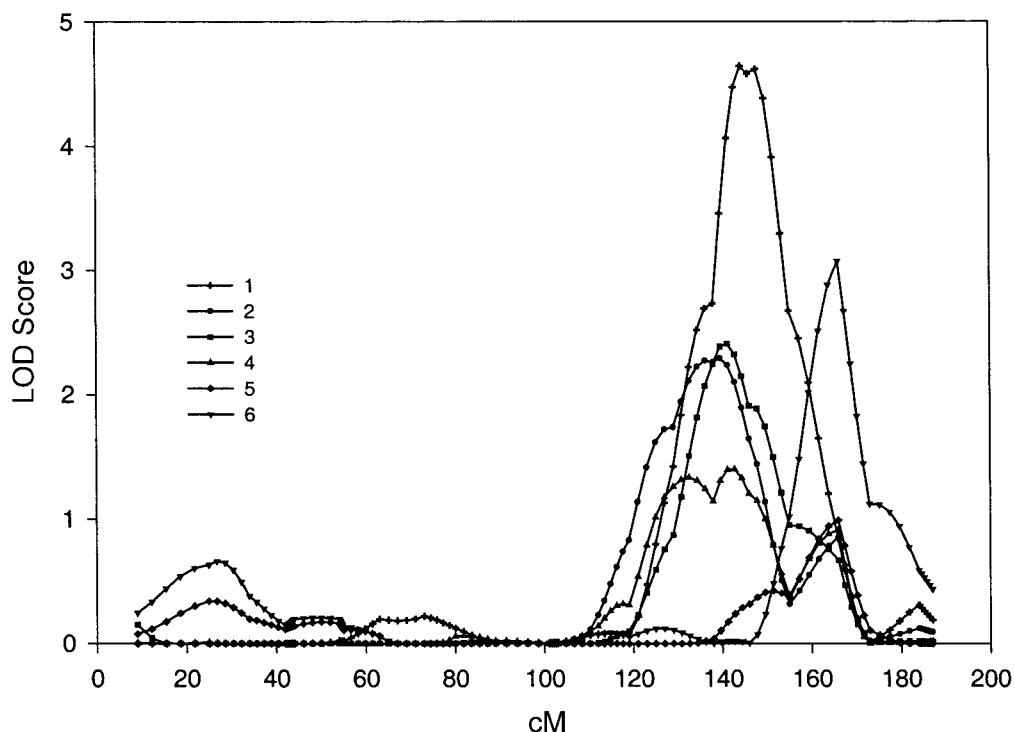
Chromosome and Data Set	Position (cM)	Nearest Marker	LOD Score
1:			
3	96.8	D1S1665	1.82
5	118.8	D1S551	1.35
6	202.0	D1S518	1.58
5	233.0	D1S2141	1.29
2:			
1	253.8	D2S2968	1.75
3:			
5	22.0	D3S1304	1.13
6	85.6	D3S4542	1.91
2	170.6	D3S1763	1.22
4	170.6	D3S1763	1.45
1	214.6	D3S2418	1.07
4	216.0	D3S2418	1.02
4:			
3	91.0	D4S2361	1.59
2	93.0	D4S2361	1.58
3	146.0	D4S1625	1.54
5	146.0	D4S1625	1.18
6	146.0	D4S1625	1.07
5:			
3	.0	D5S392	1.28
5	.0	D5S392	1.38
5	13.0	D5S2505	1.15
6:			
2	139.6	D6S1009	2.29
3	141.2	D6S1009	2.41
4	142.8	GATA184A08	1.40
1	144.4	GATA184A08	4.64
6	166.0	D6S305	3.08
8:			
2	30.4	D8S1106	2.78
3	30.4	D8S1106	2.29
4	41.2	D8S136	1.01
9:			
6	48.0	D9S1121	1.02
5	66.0	D9S301	1.09
5	89.0	D9S1120	1.70
6	89.0	D9S1120	1.25
4	89.6	D9S1120	1.69
1	104.0	D9S910	1.21
10:			
4	33.0	D10S1430	1.22
6	35.6	D10S1430	1.89
11:			
5	97.8	D11S2000	1.09
3	115.0	D11S1998	1.30
5	115.0	D11S1998	1.30
6	117.0	D11S1998	2.29
2	119.0	D11S4464	3.27
12:			
5	72.0	D12S398	1.47
4	74.0	D12S1294	1.16
13:			
5	111.0	D13S285	1.23
16:			
1	44.0	D16S403	1.40
2	65.6	D16S3396	2.07
18:			
1	7.0	D18S481	1.21
19:			
6	52.0	D19S433	1.08
21:			
6	3.0	D21S1432	1.35

1996; Pratt et al. 2000). This approach uses the genotype information at a locus to decompose the phenotypic variance into a component attributable to the locus (known as a QTL), a polygenic component, and an environmental component. The genotype information at a locus is characterized by the probability that two related individuals share zero, one, or two alleles identical by descent (IBD). Time to compute IBD probabilities grows exponentially with pedigree size in Genehunter, and nine of the largest families required a prohibitively large amount of time. Therefore, we split these pedigrees into smaller pedigrees. This small loss of genetic information should be conservative with respect to linkage results. Genehunter will compute sex-specific means and can simultaneously incorporate the effects of covariates. For all analyses presented here, we computed sex-specific means and included the effects of age and age squared in the model. We also performed a secondary analysis in which smoking status was included as a covariate.

All variance components are estimated by maximum likelihood. Linkage is tested by a likelihood ratio test in which a null hypothesis of the QTL variance component being equal to zero is compared with it being greater than zero. The resulting  $\chi^2$  statistic was converted to a traditional LOD score by dividing by  $2 \ln(10)$ . It is possible to obtain estimates of the proportion of the total phenotypic variation due to the QTL; however, two recent papers (Goring et al. 2001; Allison et al. 2002) have shown that the estimate of this effect is strongly correlated with the LOD score estimate and thus estimates the true effect poorly. Therefore, we will not present effect estimates here. Heritability estimates for all six of the BMI measures were obtained by variance components as implemented in SOLAR (Almasy and Blangero 1998). The FHS is population-based and is not selected for any particular trait; therefore, no ascertainment correction is necessary.

## Results

Table 1 displays the descriptive statistics for the variables used here in each of the six data sets. The number of individuals decreases over time, because some individuals die and others drop out of the study. The unusually large gap in average age between data sets 1 and 2 is due to the large time gap between offspring exam 1 (1971–1975) and offspring exam 2 (1979–1982). The time between the other exams is approximately four years in all cases. The average difference in ages is <4 years, since the oldest subjects are leaving the study because of death. This is reflected in the decreasing SD with age. It is well known that BMI increases with age, and this is confirmed again in the present study. The heritability of BMI increases with age, and all heritability estimates are significant at  $P < .0001$ . The observed



**Figure 1** Genome scans of BMI for all six data sets on chromosome 6. The symbols that distinguish each data set are placed at the location (in cM) where the LOD scores were actually computed.

skewness and kurtosis are not excessive (Allison et al. 1999); therefore, we did not transform the data prior to linkage analysis.

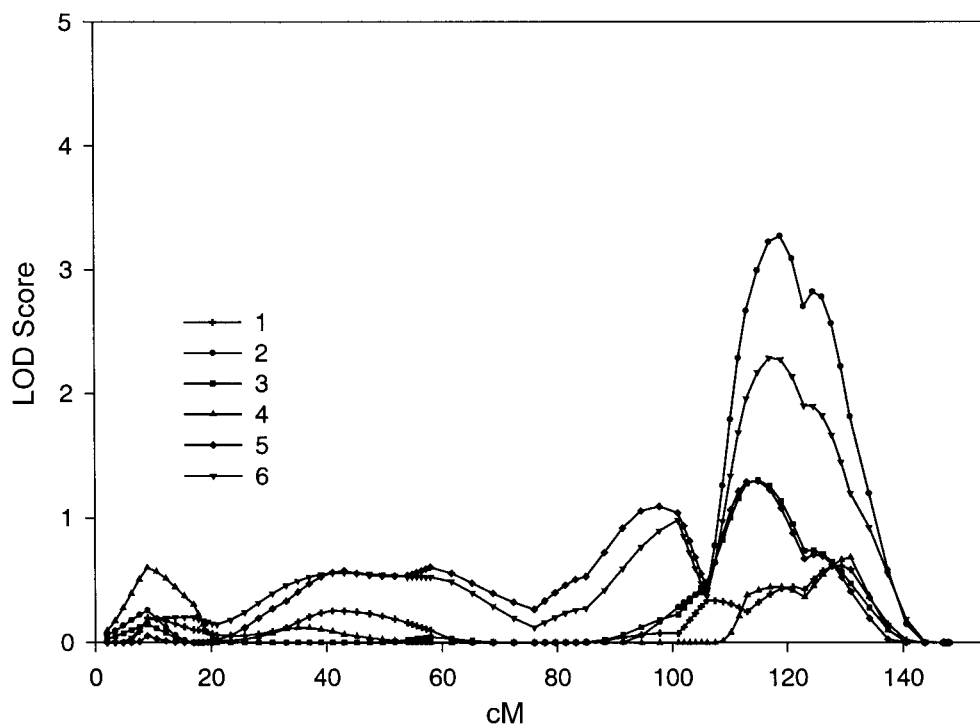
All maximum LOD scores  $>1.0$  are shown in table 2. The highest LOD score was 4.64, on chromosome 6q23-25 in data set 1. Data sets 2, 3, 4, and 6 supported linkage in this region of chromosome 6, with maximum LOD scores of 2.29, 2.41, 1.40, and 3.08, respectively. Data set 5 had a LOD score of 0.99 in the same region. This region on chromosome 6 contains the markers D6S1009, GATA184A08, D6S2436, and D6S305. Figure 1 shows all LOD scores on chromosome 6 for all six data sets graphically. There is one other chromosomal region that shows substantial evidence across multiple exams. The highest LOD score on chromosome 11q14 was 3.27, in data set 2. Data sets 3, 5, and 6 supported linkage in this region of chromosome 11, with LOD scores of 1.30, 1.30, and 2.29, respectively. Data sets 1 and 4 had LOD scores of 0.61 and 0.68 in the same region. This region of chromosome 11 contains the markers D11S1998, D11S4464, and D11S912. Figure 2 shows all LOD scores on chromosome 11 for all six data sets graphically. Chromosome 8p21-22 is the only other region in which at least two data sets had a LOD score  $>2.0$  in the same region. The region containing D8S1106, D8S1145, and D8S136 on chromosome 8 had LOD scores of 0.80, 2.78,

2.29, and 1.01, for data sets 1–4, respectively. Data sets 5 and 6 did not have a LOD score  $>0.59$  (corresponding to  $P < .05$ ) in this region. Note that, in the linkage analyses, several measures had multiple peaks on the same chromosome. This is reflected in table 2 by multiple entries for the same measure on a chromosome. In a separate analysis, we repeated the genome scan and included cigarette smoking as a covariate. The LOD scores were unchanged.

There are no obvious genetic relationships with age—that is, there are no monotonically increasing or decreasing LOD scores across all six data sets. It is interesting to note, however, that figure 1 shows two distinct peaks on chromosome 6 for all data sets, the first around 144 cM and the second around 166 cM. The maximum LOD scores around the peak at 144 cM tend to decrease across all data sets. The maximum LOD scores around the peak at 166 cM tend to increase across all data sets. This pattern suggests the possibility of two loci with opposite age effects.

## Discussion

We have found substantial evidence for linkage of BMI to two regions, on chromosomes 6q23-25 and 11q14. Both of these regions have been previously reported in



**Figure 2** Genome scans of BMI for all six data sets on chromosome 11. The symbols that distinguish each data set are placed at the location (in cM) where the LOD scores were actually computed.

the literature. For chromosome 6q23-25, Feitosa et al. (2002) found a LOD score of 1.6 in this same region in the NHLBI FHS. The review by Rankinen et al. (2002) suggests no candidate genes or other QTLs in this region. For chromosome 11q14, Hanson et al. (1998) found a LOD score of 3.6 in this region in the Pima Indians. The review by Rankinen et al. (2002) indicates a possible QTL in this region for the obesity-related traits percentage of body fat and energy expenditure.

We found some evidence for linkage of BMI to chromosome 8p21-22. This region also contains the lipoprotein lipase gene, which has been linked or associated with BMI or other obesity-related traits in multiple studies (see Rankinen et al. 2002).

A major strength of the present study is the availability of multiple measurements on the same individuals across 28 years of the FHS. The regions with the highest LOD scores had supporting evidence for linkage across all data sets. For chromosome 6, all six data sets yielded a LOD score  $>0.59$  ( $P < .05$ ) within 15 cM of the highest LOD score, 4.64. Similarly, for chromosome 11, all six data sets yielded a LOD score  $>0.59$  ( $P < .05$ ) within 10 cM of the highest LOD score, 3.27 (see figs. 1 and 2).

These data sets are not independent; each of the six data sets consists of observations on the same set of individuals in the FHS. Therefore, the common regions

of linkage that we observe do not constitute replication. However, these results do indicate that these methods are somewhat robust with respect to the measurement error and temporal changes in BMI.

To examine the overall similarity between results, we computed all possible correlations between LOD scores. The correlations ranged from 0.11 to 0.61. As expected, the highest correlation was between those data sets that were closest in time (e.g., data sets 2 and 3), and the lowest correlations were between data sets that were far apart in time (e.g., data sets 1 and 6). That the highest correlation was only 0.61 supports the hypothesis of substantial measurement error in these observations.

There is substantial variation in the maximum LOD scores for the two regions of strong linkage. A natural hypothesis is that this variation is due to the aging of the FHS participants. However, there are no obvious age-related differences—for example, monotonically increasing or decreasing LOD scores across the data sets. There is some tendency for the major and minor LOD score peaks on chromosome 6 to vary inversely with each other across age (see fig. 1). The possibility of two loci on chromosome 6 is supported by the study by Duggirala et al. (2001), who found these same two regions on chromosome 6 linked to fasting insulin levels, a trait related to obesity. Nonetheless, given the

potentially large error in location estimates (Roberts et al. 1999), the hypothesis of two loci affecting BMI on chromosome 6 should remain tentative.

The sensitivity of variance-components linkage analysis to violations of the assumption of normally distributed data is well known (Allison et al. 1999). This sensitivity can cause an error rate in excess of the nominal (usually 0.05) error rate. Allison et al. (1999) show that this sensitivity can be caused by high rates of skewness and kurtosis, small sample sizes, and large residual familial correlations. In the present study, the skewness and kurtosis, reported in table 1, are not excessive. We note that our highest LOD score occurred in the first data set, which had the smallest skewness and kurtosis. Furthermore, our data sets are all relatively large by genome-scan standards, and the residual heritability, after accounting for the effect of the QTL, is typically small ( $h^2 < 0.20$ ). In addition, both of our major linkage regions have been reported previously. Thus, there is little evidence that our results might be spurious because of violations of normality assumptions.

The present study and the recent studies by Hunt et al. (2001), Feitosa et al. (2002), Wu et al. (2002), and Zhu et al. (2002) all report linkage of BMI to chromosomal regions that have already been reported in the literature. Of the recent studies, only the study by Deng et al. (2002) reported a new region. If future genome scans of BMI also reveal substantial overlap, then it may be that the majority of QTLs affecting BMI have been, however crudely, detected and localized.

## Acknowledgments

This research was supported by the FHS, National Heart, Lung, and Blood Institute contract HC-25195.

## Electronic-Database Information

The accession number and URL for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for obesity [MIM #601665])

## References

Allison DB, Fernandez JR, Heo M, Zhu S, Etzel C, Beasley TM, Amos CI (2002) Bias in estimates of quantitative-trait locus effect in genome scans: demonstration of the phenomenon and a method-of-moments procedure for reducing bias. *Am J Hum Genet* 70:575–585

Allison DB, Neale MC, Zannolli R, Schork NJ, Amos CI,

Blangero J (1999) Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am J Hum Genet* 65:531–544

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198–1211

Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 54:535–543

Cupples LA, D'Agostino RB, Kiely D (1988) The Framingham Heart Study, section 35: an epidemiological investigation of cardiovascular disease survival following cardiovascular events: 30 year follow-up. National Heart, Lung, and Blood Institute, Bethesda, MD

Deng HW, Deng H, Liu YJ, Liu YZ, Xu FH, Shen H, Conway T, Li JL, Huang QY, Davies KM, Recker RR (2002) A genomewide linkage scan for quantitative-trait loci for obesity phenotypes. *Am J Hum Genet* 70:1138–1151

Duggirala R, Blangero J, Almasy L, Arya R, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP (2001) A major locus for fasting insulin concentrations and insulin resistance on chromosome 6q with strong pleiotropic effects on obesity-related phenotypes in nondiabetic Mexican Americans. *Am J Hum Genet* 68:1149–1164

Dyke B (1996) PEDSYS: a pedigree data management system. Southwest Foundation for Biomedical Research, San Antonio

Feitosa MF, Borecki IB, Rich SS, Arnett DK, Sholinsky P, Myers RH, Leppert M, Province MA (2002) Quantitative-trait loci influencing body-mass index reside on chromosomes 7 and 13: the National Heart, Lung, and Blood Institute Family Heart Study. *Am J Hum Genet* 70:72–82

Goring HH, Terwilliger JD, Blangero J (2001) Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 69:1357–1369

Hanson RL, Ehm MG, Pettitt DJ, Prochazka M, Thompson DB, Timberlake D, Foroud T, Kobes S, Baier L, Burns DK, Almasy L, Blangero J, Garvey WT, Bennett PH, Knowler WC (1998) An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* 63:1130–1138

Hunt SC, Abkevich V, Hensel CH, Gutin A, Neff CD, Russell DL, Tran T, Hong X, Jammulapati S, Riley R, Weaver-Feldhaus J, Macalma T, Richards MM, Gress R, Francis M, Thomas A, Frech GC, Adams TD, Shattuck D, Stone S (2001) Linkage of body mass index to chromosome 20 in Utah pedigrees. *Hum Genet* 109:279–285

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347–1363

Lindsay RS, Kobes S, Knowler WC, Bennett PH, Hanson RL (2001) Genome-wide linkage analysis assessing parent-of-origin effects in the inheritance of type 2 diabetes and BMI in Pima Indians. *Diabetes* 50:2850–2857

Pratt SC, Daly MJ, Kruglyak L (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 66:1153–1157

Rankinen T, Perusse L, Weisnagel SJ, Snyder EE, Chagnon YC, Bouchard C (2002) The human obesity gene map: the 2001 update. *Obes Res* 10:196–243

Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999) Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* 65:876–884

Wu X, Cooper RS, Borecki I, Hanis C, Bray M, Lewis CE, Zhu X, Kan D, Luke A, Curb D (2002) A combined analysis

of genomewide linkage scans for body mass index from the National Heart, Lung, and Blood Institute Family Blood Pressure Program. *Am J Hum Genet* 70:1247–1256

Zhu X, Cooper RS, Luke A, Chen G, Wu X, Kan D, Chakravarti A, Weder A (2002) A genome-wide scan for obesity in African-Americans. *Diabetes* 51:541–544