

# Genetic Variation at the *FTO* Locus Influences *RBL2* Gene Expression

Jeremy B.M. Jowett,<sup>1</sup> Joanne E. Curran,<sup>2</sup> Matthew P. Johnson,<sup>2</sup> Melanie A. Carless,<sup>2</sup> Harald H.H. Göring,<sup>2</sup> Thomas D. Dyer,<sup>2</sup> Shelley A. Cole,<sup>2</sup> Anthony G. Comuzzie,<sup>2</sup> Jean W. MacCluer,<sup>2</sup> Eric K. Moses,<sup>2</sup> and John Blangero<sup>2</sup>

**OBJECTIVE**—Genome-wide association studies that compare the statistical association between thousands of DNA variations and a human trait have detected 958 loci across 127 different diseases and traits. However, these statistical associations only provide evidence for genomic regions likely to harbor a causal gene(s) and do not directly identify such genes. We combined gene variation and expression data in a human cohort to identify causal genes.

**RESEARCH DESIGN AND METHODS**—Global gene transcription activity was obtained for each individual in a large human cohort ( $n = 1,240$ ). These quantitative transcript data were tested for correlation with genotype data generated from the same individuals to identify gene expression patterns influenced by the variants.

**RESULTS**—Variant rs8050136 lies within intron 1 of the *FTO* gene on chromosome 16 and marks a locus strongly associated with type 2 diabetes and obesity and widely replicated across many populations. We report that genetic variation at this locus does not influence *FTO* gene expression levels ( $P = 0.38$ ), but is strongly correlated with expression of *RBL2* ( $P = 2.7 \times 10^{-5}$ ), ~270,000 base pairs distant to *FTO*.

**CONCLUSIONS**—These data suggest that variants at *FTO* influence *RBL2* gene expression at large genetic distances. This observation underscores the complexity of human transcriptional regulation and highlights the utility of large human cohorts in which both genetic variation and global gene expression data are available to identify disease genes. Expedient identification of genes mediating the effects of genome-wide association study-identified loci will enable mechanism-of-action studies and accelerate understanding of human disease processes under genetic influence. *Diabetes* 59:726–732, 2010

The identification of genes contributing to disease development in humans will provide medical benefits in two ways. First, the defining of causative risk predisposing genetic variation will allow improvements to be made to current clinical risk models for disease onset and progression. Second, knowledge of the gene's identity that mediates the biological

consequences of this variation will improve understanding of the molecular pathways that participate in disease pathogenic processes. Genome-wide association studies have yielded 958 loci associated with 127 human diseases and traits over the past several years (1). Yet these associations, like many successful genetic linkage analysis studies that predate them (2), only identify a genetic region or loci and do not directly locate the causative variant(s). Without comprehensive sequence information to achieve better resolution, the associated variant may merely be in linkage disequilibrium with another untyped functional variant. Depending upon the population and sample design, the span of potential linkage disequilibrium may easily be as large as 500 kb (3) and potentially much larger (4). Alternatively, the observed associated variant may be directly functional, working to modulate transcriptional elements operating at large genetic distances. Without disease gene identifications, the benefits to medicine cannot be fully realized.

A major limitation to identification of the genes through which these loci cause disease has been the dearth of large-scale cohorts in which both gene expression and genotype data have been measured in the same individuals (5). In this study, we analyze the San Antonio Family Heart Study (SAFHS), where large-scale gene expression and variation measurements have been undertaken in the same individuals (6) and demonstrate its utility in identification of genes mediating the consequences of biological variation. We apply this approach to the *FTO* locus (Entrez GeneID: 79068) that was the most strongly associated region identified by genome-wide association study that predisposes to type 2 diabetes through an effect on body mass (7–10) and has been widely replicated (11). These data indicate that human transcriptional regulation is more complex than has been previously anticipated.

## RESEARCH DESIGN AND METHODS

**Ethics statement.** All study participants provided informed consent. The study and all protocols presented here were approved by the Institutional Review Board at the University of Texas Health Science Center at San Antonio.

**Study population.** This study was based on participants in the San Antonio Family Heart Study (12), which focuses on genetic aspects of cardiovascular health and disease in Mexican Americans.

**Expression profiling.** Total RNA was isolated from lymphocyte samples from 1,240 individuals and gene expression measured by hybridization of labeled cRNA to Illumina Sentrix Human Whole Genome (WG-6) Series I BeadChips and scanning on the Illumina BeadArray 500GX Reader as described previously (6) and in the supplementary methods section, available in an online appendix at <http://diabetes.diabetesjournals.org/content/full/db09-1277/DC1>.

**Identification of expressed transcripts.** To identify transcripts that exhibited sufficient quantitative expression in lymphocytes, the distribution of expression values for a given transcript was compared with the distribution of the expression values of the controls that are embedded in each chip. For each transcript, we performed a  $\chi^2$  “tail” test of whether there was a significant

From the <sup>1</sup>Department of Genomics and Systems Biology, Baker IDI Heart and Diabetes Institute, Melbourne, Victoria, Australia; and the <sup>2</sup>Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas.

Corresponding author: Jeremy B.M. Jowett, [jowett@bakeridi.edu.au](mailto:jowett@bakeridi.edu.au). Received 27 August 2009 and accepted 1 December 2009. Published ahead of print at <http://diabetes.diabetesjournals.org> on 15 December 2009. DOI: 10.2337/db09-1277.

© 2010 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

excess of samples with values above the 95th percentile of the control null distribution. This test was used because it allows detection of even those transcripts that are clearly present above baseline levels in only a subset of individuals, but not detectable at these levels in most individuals. Using a false discovery rate of 0.05, we identified 20,413 transcripts that exhibited significant expression by this criterion (6).

**Standardization of expression values.** To minimize the influence of overall signal levels, which may reflect RNA quantity and quality rather than a true biological difference among individuals, abundance values of all 20,413 retained transcripts were first standardized by  $z$  scoring within individuals (using decile percentage bins of transcripts, grouped by average log-transformed raw signals across individuals), followed by linear regression against individual-specific average log-transformed raw signal and its squared value. Lastly, for each transcript, we directly normalized these residual expression scores using an inverse Gaussian transformation across individuals, to ensure that the assumptions underlying variance component-based analyses were not violated. This conservative procedure results in normalized expression phenotypes that are comparable among individuals and across transcripts (6).

**Microsatellite marker genotyping.** The Human MapPairs Genome-Wide Screening Set Versions 6 and 8 from Research Genetics (Huntsville, AL) were used. Genotyping occurred according to the manufacturer's instructions by PCR on lymphocyte-derived DNA samples from study participants. The PCR products were subsequently pooled into multiplex panels for genotype calling on an automated DNA sequencer (model 377 with Genescan 672 and Genotyper software programs; Applied Biosystems, Foster City, CA). Overall, 1,345 individuals were genotyped at 432 highly polymorphic microsatellite markers, distributed with average intermarker spacing of <10 cM across all 22 autosomes. The median distance between the detected autosomal transcripts and the nearest marker is ~2.7 cM, and 80 and 90% of transcripts are within 4.3 and 5.5 cM from the nearest marker, respectively. These genotypes were subjected to extensive data cleaning that has been previously described (6).

**Single nucleotide polymorphism genotyping.** The Illumina HumanHap550 Genotyping BeadChip was used to genotype single nucleotide polymorphisms (SNPs) in 858 individuals in whom gene expression data were available using the Infinium II Assay (Illumina, San Diego, CA). The Infinium II Assay combines a single-tube, whole-genome amplification method with direct, array-based capture and enzymatic scoring of the SNP loci. Locus discrimination is provided by a combination of sequence-specific hybridization capture and array-based, single-base primer extensions. A single primer is used to interrogate a SNP locus. The 3' end of the primer is juxtaposed to the SNP site, and extension of the primer incorporates a biotin-labeled nucleotide (C or G) or a dinitrophenyl-labeled nucleotide (A or T). Signal amplification of the incorporated label further improves the overall signal-to-noise ratio of the assay. The Infinium II genotyping was performed using Illumina's standard protocols on a Tecan Freedom Evo Robot. After staining, each BeadChip was imaged on the Illumina BeadArray 500GX Reader using Illumina BeadScan image data acquisition software (Version 3.2.6). Genotype data were then assessed using the Genotyping Module of Illumina's BeadStudio software (Version 2.0).

**Genotype data cleaning.** Multivariate methods generally require that individuals with missing data be excluded from the analysis, which can result in a significant decrease in sample size. To minimize this issue, we used likelihood-based imputation with the MERLIN computer package as has been previously described (6) to rapidly impute SNP data in pedigrees. For each pedigree, imputation is performed and the posterior genotypic probabilities for each missing genotype are stored. These posterior probabilities are used to construct an appropriately weighted covariate for each SNP that is then used in association analyses.

To eliminate the potential for hidden population stratification that the standard fixed-effect association approach is susceptible to, we also used the quantitative transmission disequilibrium test approach to association testing implemented in SOLAR (13). Although this approach exhibits substantially decreased power than measured genotype analysis (14), it maintains the appropriate significance levels under the null hypothesis, even in the presence of population stratification. However, because of its poor power, we used it only as a check to identify consistency of qualitative inferences drawn from measured genotype analysis.

**Statistical genetic analysis: linkage and association.** All statistical analyses on related individuals were performed using variance component-based methodology and software package SOLAR Version 4.0 (15). As preparation for linkage analysis, the probabilities of multipoint identity-by-descent allele sharing among pairs of related individuals were computed by the Monte Carlo Markov chain multipoint approach implemented in software package Loki (16), using the genotypes at all linked markers jointly in the computations. Maximum likelihood marker allele frequencies were used as previously described (6). Covariates were included in the variance components framework as linear predictors of phenotype. Measured genotype analysis (17),

embedded in a variance component-based linkage model, was used for association testing, assuming an additive model of allelic effect (i.e., the SNP genotypes AA, AB, and BB were coded as -1, 0, and 1, respectively, and used as a linear predictor of phenotype) (13).

**Correction for multiple testing.** To correct for the effect of multiple testing for a given phenotype, we estimated the effective number of SNPs using the method of Li and Ji (18), which uses a modification of an earlier approach by Nyholt (19). After obtaining the effective number of SNPs, we used a modified Bonferroni procedure to identify a target  $\alpha$  level that would maintain an overall significance level of  $\leq 0.05$ .

**Heritability analysis.** Heritability analysis was performed under the classical approach decomposing the phenotypic variance into independent genetic and environmental components, assuming an additive model of gene action ("narrow sense" heritability) and expected kinship coefficients based on the observed intrafamilial relationships.

## RESULTS

Expression profiles were obtained from the lymphocytes of 1,240 individuals in the SAFHS (6). The potential consequences of variants genotyped in this cohort were assessed by performing correlation analysis of the genetic variation using a measured genotype approach with the high-dimensional quantitative expression data.

The obesity-associated locus within the *FTO* gene is a 47-kb region of high linkage disequilibrium (LD) that spans part of intron 1, all of exon 2, and part of intron 2 of the *FTO* gene. Two SNPs in this region, rs8050136 and rs9939609, are in high LD ( $r^2 > 0.95$ ) with each other and have been strongly associated with type 2 diabetes ( $P = 5.2 \times 10^{-8}$ ) (20) and obesity ( $P = 1.1 \times 10^{-47}$ ) (11). We examined the association of rs8050136 with levels of transcripts up to 5 Mb in either direction of the SNP. No association with *FTO* gene expression was observed ( $P = 0.54$ ;  $n = 854$ ); however, a strong association was observed with the retinoblastoma-like 2 (*RBL2*; Entrez GeneID: 5934) transcript ( $P = 1.5 \times 10^{-5}$ ;  $n = 854$ ) that is located ~270 kb proximal to *FTO* (Fig. 1).

To avoid missing the potential effects of untyped genetic variation and capture of the combined effects of all genetic variation in the region on *RBL2* and *FTO* expression, we performed multipoint genetic linkage analysis on chromosome 16 using *RBL2* and *FTO* transcript levels as expression quantitative traits. Linkage analysis was conducted on highly polymorphic simple tandem repeat markers that have been previously described in this cohort (12). We observed a nominally significant linkage signal for *RBL2* expression (logarithm of odds [LOD] score = 0.997,  $P = 0.016$ ;  $n = 1,240$ ) but not *FTO* expression directly at the *RBL2-FTO* locus (Fig. 2), indicating that only *RBL2* expression is likely to be affected by genetic variation in this region.

To identify the genetic variation in this region responsible for the observed linkage signal and to test whether rs8050136 was a contributor, we genotyped 136 SNPs spanning 706 kb in 858 of the same individuals in whom expression levels were measured. We observed strong association of several groups of variants with *RBL2* expression using the measured genotype test, one of which included rs8050136 ( $P = 1.5 \times 10^{-5}$ ; Fig. 3 and supplementary Table 1). Two other regions showed strong association, one located directly at the location of the *RBL2* structural gene itself (rs8043918;  $P = 1.8 \times 10^{-17}$ ), potentially affecting local promoter elements, and the second region 87.4 kb distal to rs8050136 at 52.5 Mb and within the *FTO* gene, but spanning introns 4 and 5 and exon 5 of *FTO* (rs7197983;  $P = 3.8 \times 10^{-8}$ ). The associations remained significant after adjustment for multiple testing after al-

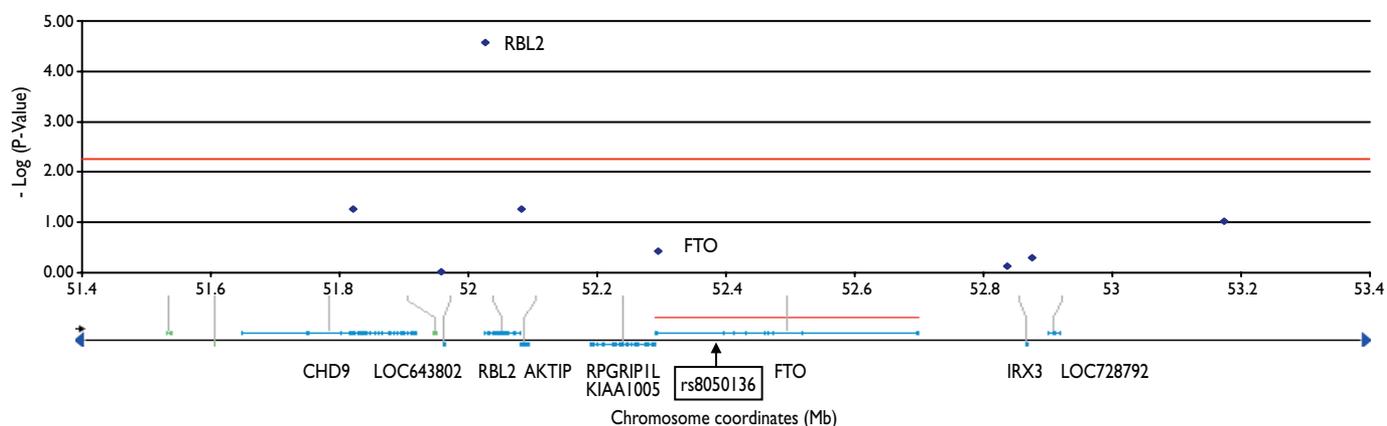


FIG. 1. Association of the rs8050136 SNP with expression levels of genes within 5 Mb of the SNP coordinate in 854 individuals. Strong association of the SNP with *RBL2* is observed, whereas no association with *FTO* is seen. The intron/exon structure of the genes in this region is illustrated by the lines and blocks beneath the x-axis (thin line, intron; block, exon). The negative log of the *P* value is plotted against the coordinates in megabase pairs (Mb). The horizontal line on the y-axis at 2.2 indicates the *P* value for experiment-wide significance following adjustment for multiple testing ( $P = 0.0063$ ).

lowing for LD (target  $\alpha P = 4.9 \times 10^{-4}$ ). To determine whether these associations were independent of background linkage, we undertook the quantitative transmission disequilibrium test, which is resistant to inflation of type I error rate in the presence of linkage (14). The following *P* values were observed for association with *RBL2*: rs8050136,  $P = 0.006$ ; rs8043918,  $P = 6.6 \times 10^{-5}$ ; and rs7197983,  $P = 0.027$ , suggesting that these associations are independent of background linkage (supplementary Table 1). The variants were also tested for association with *FTO* gene expression, but none remained significant after adjustment for multiple testing (supplementary Table 1).

The minor allele frequency (MAF) of rs8050136 (A allele) was 0.23 in the SAFHS, consistent with prior data in the Mexican American population (MAF = 0.20) (21) and lower than that observed in the CEPH (Utah residents with ancestry from northern and western Europe) cohort (MAF = 0.46) and Yoruba in Ibadan, Nigeria (MAF = 0.46), but higher than the Han Chinese in Beijing, China (MAF = 0.14), and Japanese in Tokyo cohorts (MAF = 0.18). We

observed a similar 40-kb region of high LD ( $r^2 > 0.5$ ) around rs8050136 and rs9939609 that rapidly fell outside this interval as previously reported (10).

We performed conditional linkage analysis to determine whether rs8050136 contributed to the observed linkage signal for *RBL2* gene expression. Incorporation of rs8050136 into the linkage model as a covariate reduced the LOD score from 0.997 to 0.201 (Fig. 4B). Further addition of the two SNPs most strongly associated with *RBL2* expression levels (rs8043918 and rs9921587;  $n = 852$ ) reduced the LOD score to zero (Fig. 4C), indicating that either these variants or those correlated with them account for the majority of the observed linkage signal, although the contribution of additional variants of small effect size cannot be ruled out.

The initial observation of association of rs8050136 and rs9939609 was with type 2 diabetes; however, after adjustment for BMI, the association was abrogated (10). We tested for association of rs8050136 with type 2 diabetes in the SAFHS and observed a nominal association ( $P = 0.050$ ). Furthermore, and consistent with its initial discov-

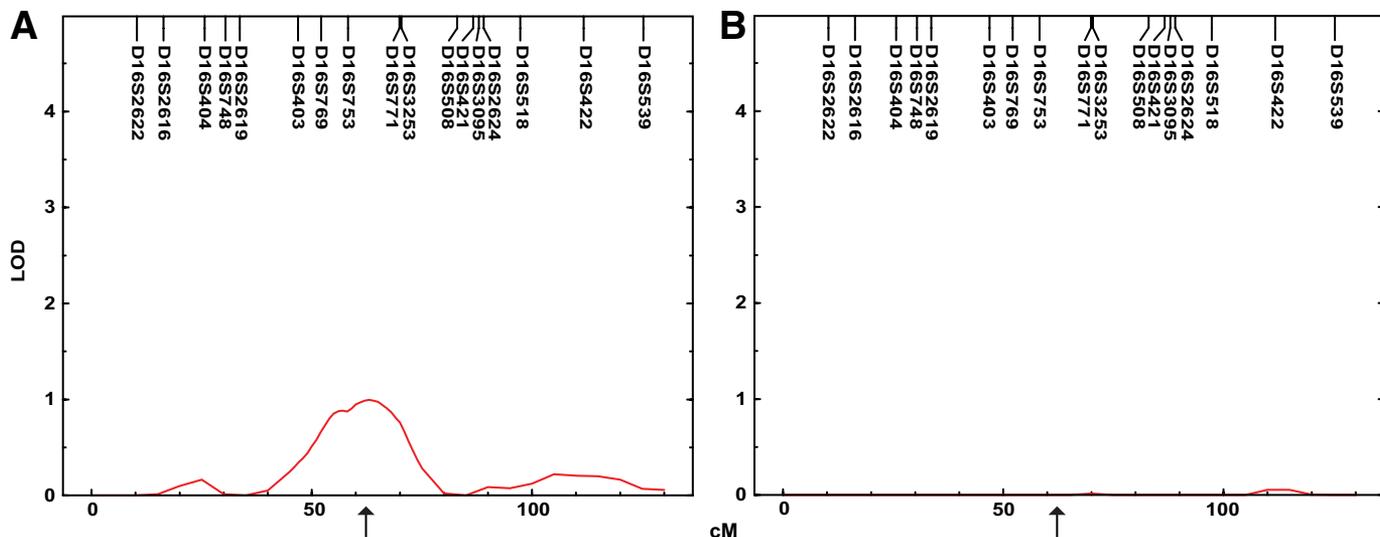
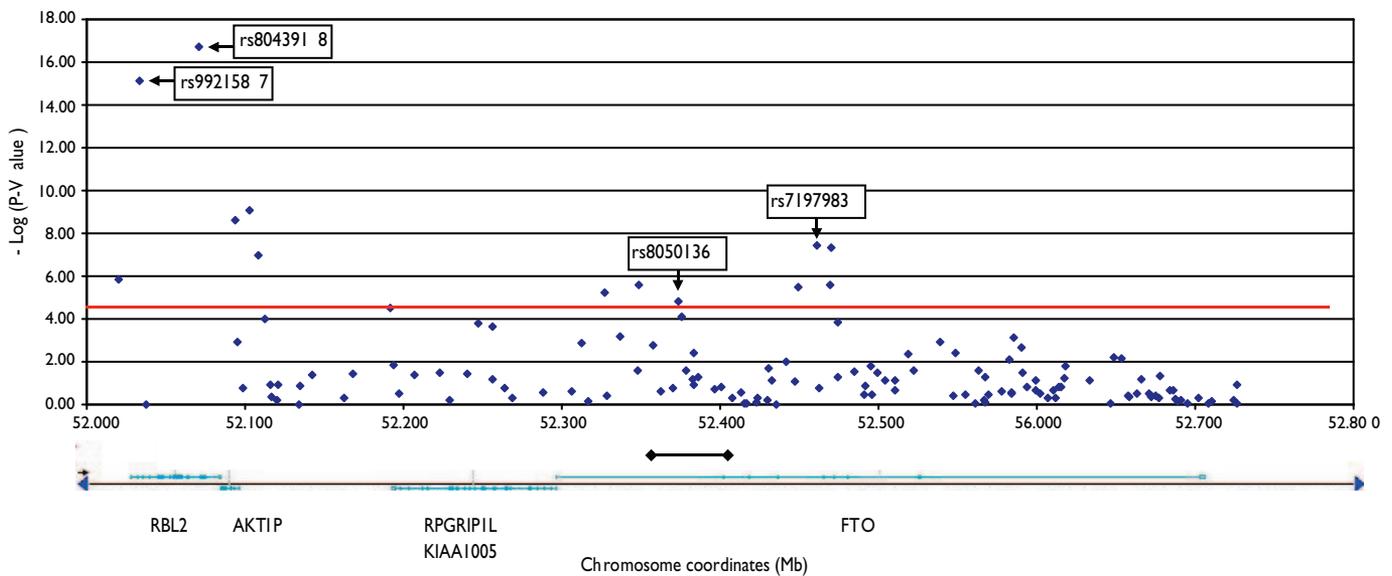


FIG. 2. Multipoint genetic linkage analysis on chromosome 16 for *RBL2* and *FTO* gene expression. Genome-wide linkage analysis for *RBL2* expression (A) and *FTO* expression (B) is shown for 1,240 individuals and reveals a linkage signal on chromosome 16 for *RBL2* but not for *FTO*. The LOD score is indicated on the y-axis, and the genetic coordinates are shown on the x-axis in centimorgans (cM). The black arrows indicate the structural position of the *RBL2* and *FTO* genes.



**FIG. 3.** Association of 136 SNPs with *RBL2* gene expression levels across the *RBL2-FTO* gene region in 858 individuals. Three regions of significant association were observed. The first, near *RBL2*, is denoted by rs8043918; the second is denoted by the position of rs8050136; and the third spanning introns 4 and 5 and exon 5 of *FTO* is denoted by rs7197983. The horizontal bar beneath the x-axis at 52.4 Mb indicates the region of high LD within *FTO* where rs8050136 is located. The horizontal line on the y-axis at 4.5 indicates the experiment-wide significance level after adjustment for multiple testing ( $P = 4.9 \times 10^{-4}$ ). The lines and blocks beneath the x-axis indicate the intron and exon boundaries of the genes, respectively. The negative log of the  $P$  value of the association is indicated on the y-axis; the genetic distance in megabase pairs (Mb) is shown on the x-axis.

ery, the association was lost after adjustment for BMI ( $P = 0.20$ ). The strength of this association is lower than that previously reported, likely due to the smaller sample size ( $n = 854$ ) in the SAFHS cohort; however, the observed trend is consistent with prior published data. Although we observed no association between *FTO* and *RBL2* expression levels with diabetes traits, the SNP most strongly associated with *RBL2* expression (rs8043918) exhibited nominal association with diabetes status ( $P = 0.047$ ) and homeostasis model assessment–insulin resistance (HOMA-IR) ( $P = 0.028$ ) after adjustment for BMI.

## DISCUSSION

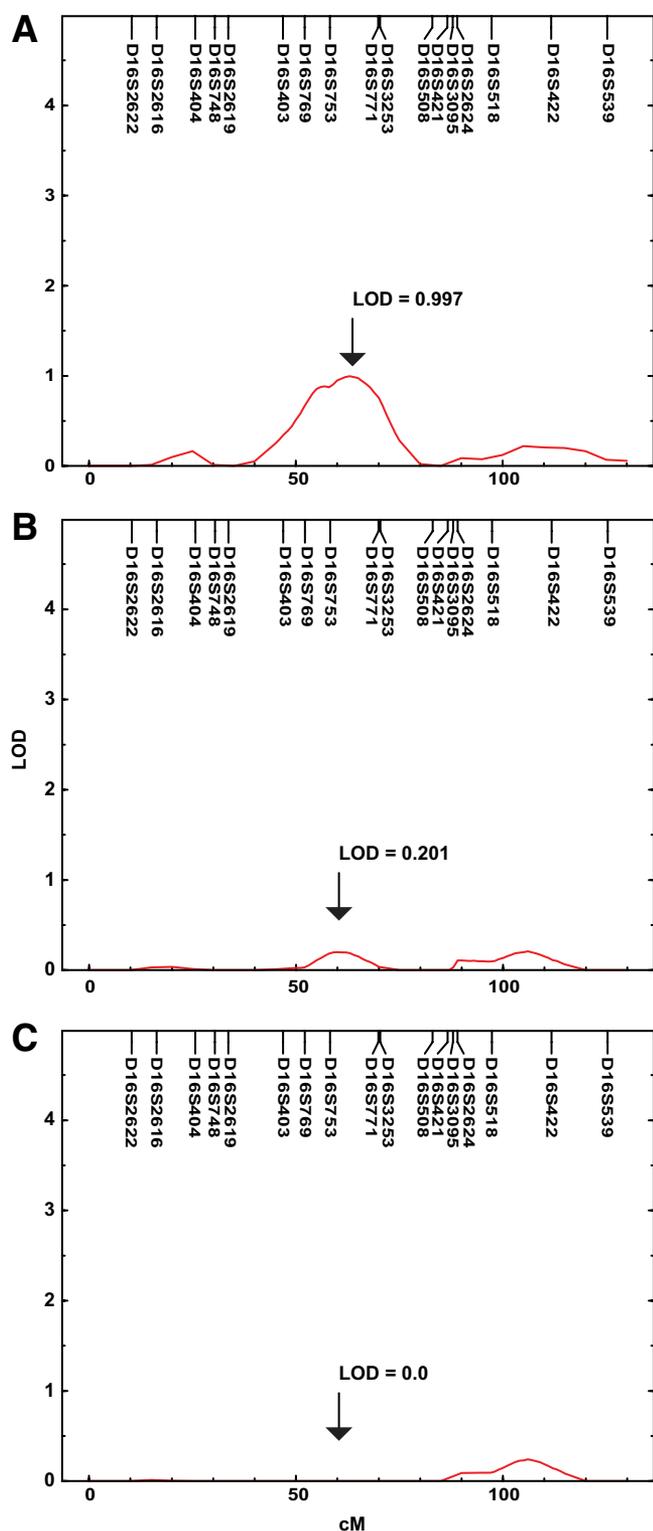
The majority of the replicated genome-wide association studies have identified variants that lie in intronic or intergenic regions; however, follow-up on identification of genes mediating the consequence of biological variation has been problematic because candidate genes may lie hundreds of kilobases from the variant (1). Mapping expression quantitative trait loci have the potential to identify which genes mediate the effects of genetic variation. We used a large family-based cohort where both genotype and gene expression data are available from the same individuals to address the question of which genes juxtaposed to disease-associated loci confer the biological consequences of the variation. We examined the most strongly type 2 diabetes- and obesity-associated loci denoted by rs8050136 located in intron 1 of the *FTO* gene. We observed no effect of this variant on *FTO* gene expression levels, but observed a strong association with *RBL2* gene expression located ~270-kb proximal to *FTO*. Genetic linkage analysis in the same cohort revealed a modest linkage signal for *RBL2* expression and none for *FTO*. Variants typed in this region, including rs8050136, were significantly associated with *RBL2* expression and found to account for the majority of the linkage signal, whereas no variants were associated with *FTO* expression after adjustment for multiple testing. These data implicate

*RBL2* as the major contributor to the biological consequences of rs8050136 or genetic variation in LD with it.

The SNPs rs8050136 and rs9939609 that are in high LD ( $r^2 > 0.95$ ) have been genotyped in an extensive number of replication studies that predominantly confirm the original findings. Because these replication studies evaluate only disease association with genotype data, our results are not in conflict with this body of literature.

The lack of correlation of SNPs rs8050136 and rs9939609 with *FTO* mRNA expression has been previously reported in human muscle and adipose tissues in a number of studies. Our data are consistent with these reports. Despite the lack of correlation with these SNPs, some human studies have been able to identify relationships between *FTO* mRNA and measures of obesity. In a large cohort of female Scandinavians ( $n = 306$ ), *FTO* mRNA was slightly higher in subcutaneous adipose tissue of obese (BMI  $>30$  kg/m<sup>2</sup>) than in nonobese (BMI  $<25$  kg/m<sup>2</sup>) individuals (22). A second large study of Danish twins ( $n = 496$ ) similarly observed increased *FTO* mRNA with increased BMI (23), as did a smaller Spanish cohort (24). In contrast, however, a study reported decreased mRNA with increased BMI in a Scandinavian cohort (25). Mouse models of obesity show decreased *FTO* mRNA expression with increased obesity (26), but conflict with the study by Church et al. (27) on an *N*-ethyl-*N*-nitrosurea (ENU)-induced COOH-terminal mutant of *FTO* (mFTO<sup>I367F</sup>) in mice that exhibited reduced levels of *FTO* protein and decreased body and fat mass. Collectively, it appears that *FTO* may play a role in obesity, although the mechanism remains controversial. In humans, because the SNPs rs8050136 and rs9939609 consistently fail to correlate with *FTO* mRNA expression across multiple tissues, it is plausible that other genes such as the neighboring *RBL2* may participate in mediating these effects and that downstream processes may have the potential to affect *FTO* gene expression indirectly.

We observed nominal association between rs8043918



**FIG. 4.** Conditional genetic linkage analysis of *RBL2* gene expression levels on chromosome 16q12. **A:** Original linkage for *RBL2* expression. **B:** Addition of rs8050136 as covariate in linkage model. **C:** Addition of rs8050136, rs8043918, and rs9921587 as covariates in linkage model in 852 individuals. LOD score is indicated for the peak linkage signal under each of the three models. Black arrow denotes the position of the *RBL2* and *FTO* genes. The *y*-axis is the LOD score; the *y*-axis is the genomic location in centimorgans (cM). Position of the simple tandem repeat markers is shown across the top border of each panel.

and HOMA-IR ( $P = 0.028$ ) and diabetes status ( $P = 0.047$ ) after adjustment for BMI, suggestive of a potential role that *RBL2* may play in obesity and diabetes. However, the observed strength of the association prior to adjustment for multiple testing indicates that the relationship should be considered preliminary at this stage. Consistent with this result, we note that in the original genome-wide association study, genetic variants juxtaposed to *RBL2* were nominally associated with type 2 diabetes ( $P \sim 1 \times 10^{-4}$ , 20). Differences in strength of these associations might be explained by different tiers of transcriptional control that operate independently and in different circumstances according to availability of transcription factors and temporal development. For example, the locus identified within intron 1 of *FTO*, although 270 kb distant, may play a specific role in control of *RBL2* gene expression in a particular tissue type and/or in a specific stage of development that influences obesity. Conversely, genetic variation directly at the *RBL2* locus potentially mediates their effects through interaction with an alternate set of transcription factors responding independently to internal programs and external stimuli and affect cellular properties that have little impact on development of obesity.

A limitation of this study is its reliance on expression from a single human tissue and currently a lack of knowledge regarding coordinate gene expression control among multiple tissues. It is unclear whether gene expression patterns observed in lymphocytes can be extrapolated to other tissues such as adipose and muscle; however, in prior work, valid inferences have been made by such extrapolations, particularly in the area of central nervous system disorders (28). Recent studies suggest, however, that variants affecting gene expression in a consistent manner among fibroblasts, B-cells, and T-cells are limited to ~17% of genes for two or more tissue types (29). Additionally, work by Heintzman et al. (30) suggest that consistency among cell types is dependent upon the type of expression control region, with promoters and insulators showing highest between-cell type consistency and enhancer regions showing greatest divergence. Extrapolating these findings to the *FTO-RBL2* locus is difficult because the control elements for the locus have not been fully delineated. Our data and that of others show that the lack of correlation between *FTO* SNPs and *FTO* gene expression was consistently observed across human muscle, adipose, and lymphocyte tissue types (22–26,31), suggesting this control region may be among those conserved between cell types. Further investigation into genome-wide transcriptional analysis across tissues is warranted in this area, particularly in metabolic disease given the potential benefits to understanding of disease processes and development of biomarkers. The SNP rs8050136 creates a cut-like 1 (*CUX1*) transcription factor binding site (26) that may participate in mediating its biological effects; however, we note that *CUX1* is broadly expressed across human tissue types including adipose, liver, muscle, and lymphoid (32).

Replication of expression datasets is an important consideration but difficult to achieve. Although large cohorts exist for replication of genetic variation analyses such as genome-wide association studies, there are far fewer large-scale cohorts that contain genome-wide expression data, making replication of expression-related findings problematic. This is further complicated by differences in tissue types that are relevant to a specific disease under investigation. As noted above, some genes share coordi-

nate expression patterns between tissues, whereas others do not. Availability of transcription factors in different cell types as well as methylation patterns contribute to this diversity. To address this, we and colleagues are seeking to establish large-scale gene expression replication cohorts with a range of tissues for analysis, which in the future may allow expression replication studies to be undertaken.

*RBL2* is a member of the retinoblastoma (RB) family of tumor suppressor genes. Like RB, *RBL2* binds members of the DNA binding E2F transcription factor family that regulate transcription of many genes associated with cell cycle progression and cellular differentiation (33). Notably, *RBL2* plays a role in preadipocyte proliferation and differentiation, specifically binding E2F in the preadipocyte quiescent nonproliferating and differentiated states (29).

Adipocyte number is a major determinant of fat mass in adults, and a recent study has shown that the number, which is set in childhood and adolescence, remains relatively constant over adult life (34). The A allele of rs8050136 (which is in high LD with the A allele of rs9939609) confers increased risk of obesity that is present by the early age of 11 years (10). We observed that the A allele of rs8050136 is strongly positively correlated with *RBL2* expression levels, raising the possibility that increased levels of *RBL2* may serve to limit clonal expansion of the preadipocyte population during development. A finite pool of adipocytes, with the environmental trigger of excess consumption of highly processed energy-dense food, may undergo cellular hypertrophy that is characterized by a hyperlipolytic state, resistance to insulin's antilipolytic effects, and increased nonesterified fatty acid flux (35). The effects of excess nonesterified fatty acids on the liver and periphery promote insulin resistance, loss of glucose homeostasis, and development of type 2 diabetes, providing a possible explanation as to why this locus was originally identified based on association with type 2 diabetes, which was dependent upon the presence of obesity.

In summary, we present data to suggest that the *RBL2* gene may participate in mediating the biological consequences of genetic variation in the fat mass and obesity-associated locus within the *FTO* gene denoted by rs8050136.

#### ACKNOWLEDGMENTS

A generous donation from the Azar/Shepperd families paid for the transcriptional profiling. The statistical genetics computer package, SOLAR, was supported by a grant from the National Institute of Mental Health (MH059490). The supercomputing facilities used for this work at the AT&T Genetics Computing Center were supported in part by a gift from the AT&T Foundation. The laboratory work was conducted in facilities constructed with support from the National Center for Research Resources (RR013556). Data collection was supported by a grant from the National Heart, Lung, and Blood Institute (HL045222).

Additional funds for transcriptional profiling, sequencing, genotyping, and statistical analysis were provided by ChemGenex Pharmaceuticals. No other potential conflicts of interest relevant to this article were reported.

We are grateful to the participants in the San Antonio Family Heart Study.

#### REFERENCES

- Hindorf LA, Junkins HA, Mehta JP, Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available from <http://www.genome.gov/26525384>. Accessed 20 December 2008
- Blangero J. Localization and identification of human quantitative trait loci: king harvest has surely come. *Curr Opin Genet Dev* 2004;14:233–240
- Collins A, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms [see comments]. *Proc Natl Acad Sci U S A* 1999;96:15173–15177
- Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, Persico I, Maestrale G, Hastie ND, Pirastu M, Visscher PM. Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Hum Mol Genet* 2004;13:25–33
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;9:356–369
- Göring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* 2007;39:1208–1216
- Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Boström KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jørgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjögren M, Steinthorsdóttir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdóttir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Wellcome Trust Case Control Consortium, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40:638–645
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;316:1336–1341
- Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 2007;8:657–662
- Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;316:889–894
- Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdóttir V, Sulem P, Helgadóttir A, Styrkarsdóttir U, Gretarsdóttir S, Thorlacius S, Jonsdóttir I, Jonsdóttir T, Olafsdóttir EJ, Olafsdóttir GH, Jonsson T, Jonsson F, Borch-Johnsen K, Hansen T, Andersen G, Jørgensen T, Lauritzen T, Aben KK, Verbeek AL, Roeleveld N, Kampman E, Yanek LR, Becker LC, Tryggvadóttir L, Rafnar T, Becker DM, Gulcher J, Kiemeneý LA, Pedersen O, Kong A, Thorsteinsdóttir U, Stefansson K. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009;41:18–24
- Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel RD, Sharp RM, Comuzzie AG, VandeBerg JL, Stern MP, MacCluer JW. Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans: the San Antonio Family Heart Study. *Circulation* 1996;94:2159–2170
- Blangero J, Göring HH, Kent J, Williams JT, Peterson CP, Almasy L, Dyer TD. Quantitative trait nucleotide analysis using bayesian model selection. *Hum Biol* 2005;77:541–549
- Havill LM, Dyer TD, Richardson DK, Mahaney MC, Blangero J. The

- quantitative trait linkage disequilibrium test: a more powerful alternative to the quantitative transmission disequilibrium test for use in the absence of population stratification. *BMC Genet* 2005;6(Suppl. 1):S91
15. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;62:1198–1211
  16. Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 1997;61:748–760
  17. Boerwinkle E, Chakraborty R, Sing CF. The use of measured genotype information in the analysis of quantitative phenotypes in man. I: models and analytical methods. *Ann Intern Med* 1986;50:181–194
  18. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005;95:221–227
  19. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004;74:765–769
  20. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678
  21. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–1320
  22. Wåhlén K, Sjölin E, Hoffstedt J. The common rs9939609 gene variant of the fat mass- and obesity-associated gene FTO is related to fat cell lipolysis. *J Lipid Res* 2008;49:607–611
  23. Grunnet LG, Nilsson E, Ling C, Hansen T, Pedersen O, Groop L, Vaag A, Poulsen P. Regulation and function of FTO mRNA expression in human skeletal muscle and subcutaneous adipose tissue. *Diabetes* 2009;58:2402–2408
  24. Zabena C, González-Sánchez JL, Martínez-Larrad MT, Torres-García A, Alvarez-Fernández-Represa J, Corbatón-Anchuelo A, Pérez-Barba M, Serrano-Ríos M. The FTO obesity gene: genotyping and gene expression analysis in morbidly obese patients. *Obes Surg* 2009;19:87–95
  25. Klötting N, Schleinitz D, Ruschke K, Berndt J, Fasshauer M, Tönjes A, Schön MR, Kovacs P, Stumvoll M, Blüher M. Inverse relationship between obesity and FTO gene expression in visceral adipose tissue in humans. *Diabetologia* 2008;51:641–647
  26. Stratigopoulos G, Padilla SL, LeDuc CA, Watson E, Hattersley AT, McCarthy MI, Zeltser LM, Chung WK, Leibel RL. Regulation of Fto/Ftm gene expression in mice and humans. *Am J Physiol Regul Integr Comp Physiol* 2008;294:R1185–R1196
  27. Church C, Lee S, Bagg EA, McTaggart JS, Deacon R, Gerken T, Lee A, Moir L, Mecinovic J, Quwallid MM, Schofield CJ, Ashcroft FM, Cox RD. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet* 2009;5:e1000599
  28. Borovecki F, Lovrecic L, Zhou J, Jeong H, Then F, Rosas HD, Hersch SM, Hogarth P, Bouzou B, Jensen RV, Krainc D. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. *Proc Natl Acad Sci U S A* 2005;102:11023–11028
  29. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009;325:1246–1250
  30. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 2009;459:108–112
  31. Grunnet LG, Brøns C, Jacobsen S, Nilsson E, Astrup A, Hansen T, Pedersen O, Poulsen P, Quistorff B, Vaag A. Increased recovery rates of phosphocreatine and inorganic phosphate after isometric contraction in oxidative muscle fibers and elevated hepatic insulin resistance in homozygous carriers of the A-allele of FTO rs9939609. *J Clin Endocrinol Metab* 2009;94:596–602
  32. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R. NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acid Res* 2007;35:D760–D765
  33. Shin EK, Shin A, Paulding C, Schaffhausen B, Yee AS. Multiple change in E2F function and regulation occur upon muscle differentiation. *Mol Cell Biol* 1995;15:2252–2262
  34. Spalding KL, Arner E, Westermark PO, Bernard S, Buchholz BA, Bergmann O, Blomqvist L, Hoffstedt J, Näslund E, Britton T, Concha H, Hassan M, Rydén M, Frisén J, Arner P. Dynamics of fat cell turnover in humans. *Nature* 2008;453:783–787
  35. Després JP, Lemieux I. Abdominal obesity and metabolic syndrome. *Nature* 2006;444:881–887