Published in final edited form as:

Metabolism. 2012 August ; 61(8): 1129–1141. doi:10.1016/j.metabol.2012.01.008.

Obesity–insulin targeted genes in the 3p26-25 region in human studies and LG/J and SM/J mice

Aldi T. Kraja^{a,*}, Heather A. Lawson^{b,*}, Donna K. Arnett^c, Ingrid B. Borecki^a, Ulrich Broeckel^d, Lisa de las Fuentes^e, Steven C. Hunt^f, Michael A. Province^a, James Cheverud^b, and D.C. Rao^g

Aldi T. Kraja: aldi@wustl.edu; Heather A. Lawson: lawsonh@pcg4.wustl.edu; Donna K. Arnett: arnett@ms.soph.uab.edu; Ingrid B. Borecki: iborecki@wustl.edu; Ulrich Broeckel: broeckel@mcw.edu; Lisa de las Fuentes: lfuentes@wustl.edu; Steven C. Hunt: steve.hunt@utah.edu; Michael A. Province: mprovince@wustl.edu; James Cheverud: cheverud@pcg.wustl.edu; D.C. Rao: rao@wubios.wustl.edu

^aDivision of Statistical Genomics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110, USA

^bDepartment of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO 63110, USA

^cDepartment of Epidemiology, University of Alabama, Birmingham, AL 35294, USA

^dIndividualized Medicine Institute, Medical College of Wisconsin, WI 53226, USA

^eCardiovascular Division Department of Medicine, Cardiovascular Imaging and Clinical Research Core Laboratory, Washington University School of Medicine 63110, St. Louis, MO, USA

^fCardiovascular Genetics Division, Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA

^gDivision of Biostatistics, Washington University School of Medicine 63110, St. Louis, MO, USA

Abstract

Identifying metabolic syndrome (MetS) genes is important for novel drug development and health care. This study extends the findings on human chromosome 3p26-25 for an identified *obesity–insulin* factor QTL, with an LOD score above 3. A focused association analysis comprising up to 9578 African American and Caucasian subjects from the HyperGEN Network (908 African Americans and 1025 whites), the Family Heart Study (3035 whites in time 1 and 1943 in time 2), and the Framingham Heart Study (1317 in Offspring and 1320 in Generation 3) was performed. The homologous mouse region was explored in an F₁₆ generation of an advanced intercross between the LG/J and SM/J inbred strains, in an experiment where 1002 animals were fed low-fat (247 males; 254 females) or high-fat (253 males; 248 females) diets. Association results in humans indicate pleiotropic effects for SNPs within or surrounding *CNTN4* on obesity, lipids and blood pressure traits and for SNPs near *IL5RA*, *TRNT1*, *CRBN*, and *LRRN1* on central obesity and blood pressure. Linkage analyses of this region in LG/J × SM/J mice identify a highly significant pleiotropic QTL peak for insulin and glucose levels, as well as response to glucose

© 2012 Elsevier Inc. All rights reserved.

*Corresponding authors. Aldi Kraja, is to be contacted at Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO 63110 USA. Heather Lawson, Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO 63110, USA.

Author contributions: ATK and HAL performed human and rodent research respectively and wrote the manuscript. ATK, HAL, DKA, IBB, UB, LdIF, SCH, MAP, JC, DCR participated in discussions and revisions of the full manuscript.

Conflicts of interest

None.

challenge. The mouse results show that insulin and glucose levels interact with high and low fat diets and differential gene expression was identified for *Crbn* and *Arl8b*. In humans, *ARL8B* resides ~137 kbps away from *BHLHE40*, expression of which shows up-regulation in response to insulin treatment. This focused human genetic analysis, incorporating mouse research evidenced that 3p26-25 has important genetic contributions to MetS components. Several of the candidate genes have functions in the brain. Their interaction with MetS and the brain warrants further investigation.

1. Introduction

Metabolic syndrome (MetS) refers to the co-occurrence of conditions involving high levels of cholesterol, obesity, impaired fasting glucose, and/or high blood pressure associated with increased risk of coronary artery disease, stroke and type 2 diabetes mellitus (T2D). According to the American Heart Association and the National Heart, Lung, and Blood Institute, MetS is present if an individual has three or more of the following components: large waist circumference (length around the waist): for men—102 cm or more, women—88 cm or more; low HDL cholesterol: men—under 40 mg/dl, women—under 50 mg/dl; triglycerides equal to or higher than 150 mg/dl; blood pressure equal to or higher than 130/85 mm Hg; fasting blood sugar (glucose) equal to or higher than 110 mg/dl. Age adjusted MetS prevalence among U.S. adults aged >20 years has increased from 29% in 1988–1994 to 34% in 1999–2006 [1]. Such a rise represents a substantial social and economic burden, with \$17 billion annual losses in productivity attributable to MetS [2]. Increase in MetS prevalence is hypothesized to result from excessive caloric intake in conjunction with increasingly sedentary lifestyles, with accompanying metabolic disruption and fat accumulation [3,4]. Familial aggregation of MetS components indicates an important genetic contribution to inter-individual variation in MetS [5]. Thus, there is much interest in identifying genetic variants associated with the susceptibility to MetS and its components.

It has been proposed that focusing on QTL regions may help offsetting much of the statistical burden of correcting for genome-wide multiple tests [6]. A 2009 survey studied the agreement among 375 genome-wide linkage analyses and 102 genome-wide association scans (GWAS). They defined ‘convergence’ as the proportion of linkage intervals for a phenotype with at least one statistically significant SNP falling within the linkage area [7]. Only two of 19 traits studied converged significantly higher than expected by chance, but when the survey was concentrated only on large studies, statistical convergence was found in 6 traits, including a BMI–obesity phenotype. To affirm cogency of the above, we investigate the agreement among linkage, association, and gene expression results by focusing on a single quantitative trait locus (QTL) region, thus limiting the number of tests performed to a local neighborhood.

The focal QTL was previously identified for an *obesity–insulin* factor mapping to chromosome 3p26-25 in Caucasian families participating in the HyperGEN study [8]. Fig. 1A summarizes the HyperGEN QTL for this latent factor with an LOD score above 3. The latent factors of the HyperGEN Caucasian cohort were produced by multivariate factor analysis with substantial contributions of several risk variables for MetS (see Methods). SNP associations with MetS components were interrogated in both African American and Caucasian cohorts of the HyperGEN study and in Caucasian cohorts of the Framingham Heart Study (Framingham HS) and the Family Heart Study (Family HS). Furthermore, in this article is incorporated information on a QTL for serum insulin and glucose levels and response to a glucose challenge in an F₁₆ advanced intercross between the LG/J and SM/J inbred mouse strains. The mouse QTL covers a region homologous to the human 3p26-25 QTL [9]. Mouse models have been used extensively to understand the genetic architecture of metabolic syndrome components and human–mouse homology is well-defined [10]. QTL

identified in mice can be used to leverage associations found in human studies [11]. The genetic variability between the LG/J and SM/J inbred mouse strains makes their crosses particularly useful as models for studying components of MetS [12]. In this study, promising candidate loci with pleiotropic effects on MetS risk factors were identified by combining QTL, SNP, expression data, and database searches from multiple human and mouse sources.

2. Methods

2.1. Human samples and phenotypes

The HyperGEN study is one of four Family Blood Pressure Program networks supported by the NHLBI to identify genetic causes of high blood pressure and hypertension [13]. In the first phase of the HyperGEN Study, hypertensive (systolic blood pressure (SBP) ≥ 140 mm Hg or diastolic blood pressure (DBP) ≥ 90 mm Hg or receiving anti-hypertensive medications) sibships eligible for recruitment consisted of probands, with onset of hypertension by age 60 and one or more hypertensive siblings. In the second phase, offspring of the hypertensive siblings were recruited. The HyperGEN Study recruited African American subjects from centers located in Birmingham, AL, and Forsyth County, NC; and Caucasians from centers in Forsyth County, NC, in Framingham, MA, in Minneapolis, MN, and in Salt Lake City, UT. Study details are described elsewhere [8]. In the Family HS [14] probands were 45–69 years old and selected from three population-based epidemiologic studies, the Framingham HS, the Atherosclerosis Risk in Communities Study (ARIC), and the Utah Health Family Tree Study (UHFTS) in Salt Lake City, Utah. The Family HS had a clinical visit in 1994–1996 (time 1), and a follow-up for a relatively smaller sample in 2002–2003 (time 2). In this analysis we used both time 1 and time 2 measurements, explained in detail elsewhere [15]. For the association analysis also included were the Framingham HS data. The Framingham HS is an observational, prospective study of risk factors for cardiovascular disease that began in 1948. In 1971 and 1975 Framingham HS recruited offspring with one parent at higher risk of cardiovascular disease (CVD) due to higher lipid levels. In 2002 and 2005 a third generation was recruited. In this analysis the Framingham HS utilized data were dbGaP originated, with measurements from visit 7 of offspring and visit 1 of generation 3, as detailed elsewhere [16,17].

Phenotypes used in these analyses are considered as risk factors for MetS and include 11 traits: body mass index (BMI), waist circumference (WC), waist-to-hip ratio (WHR), percent body fat (PBF), fasting insulin (INS), fasting glucose (GLUC), fasting triglycerides (TG), high density lipoprotein cholesterol (HDL), low density lipoprotein cholesterol (LDL), SBP and DBP. Family HS time 1 does not include PBF and Family HS time 2 does not include INS. Each variable was checked for normal distribution and transformed if needed (as for example TG to a natural logarithm TG). If a subject did not conform with 12 h of fasting then his/her INS, GLUC and TG were set to missing and the full subject's record was excluded, because factor analysis does not handle data with any missing values in the full record of a subject. Variables were adjusted for age, age², and study center within sex. In the HyperGEN African Americans, we used HaploView software [18] to identify genome wide tag SNPs. Using tag SNPs as LD independent markers and Eigenstrat software [19] on unrelated subjects, 10 eigen-vectors (principal components or PCs) were produced. This Eigenstrat model was then extended to the full set of HyperGEN African American data. The 10 PCs were included in the statistical model of associations as a means to correct for population stratification. Additionally, phenotype data had to match with accompanying genotypes of individuals in order to contribute to the association tests. As a result phenotypes and genotypes of about 9578 subjects were included in the association analyses (908 African Americans and 1025 white subjects of HyperGEN Network, 3035 whites in time 1 and 1943 in time 2 of the Family HS, and 1317 in Offspring and 1320 in Generation 3 white subjects of the Framingham HS).

Varying genotyping densities were used in these studies reflecting rapidly changing genotyping costs and technology. Independent panels of more than 370 autosomal microsatellites for the HyperGEN African American and Caucasians were used in the linkage analysis; an Affymetrix Genome-Wide Human SNP Array 6.0 with approximately 900K SNPs was used for the HyperGEN African Americans genotyping, an Affymetrix 5.0 with about 500K SNP array was used for the HyperGEN Caucasians; three Illumina chip types with 550K, 610K and 1M SNPs were used for the Family HS. In the Family HS ~2.5 M imputed SNPs were produced based on CEU HapMap haplotypes (release 22, build 36) with 100 Markov Chain Monte Carlo iterations using MACH v 1.0.16 software [20]; and Affymetrix 5.0 with 500K SNP array and a 50K SNP gene centric human focused chip were used for the Framingham HS.

2.2. Statistical methods

Multivariate analysis was used to produce factor scores that served as response variables in linkage and association analyses. The purpose was to reduce the number of independent traits tested while attempting to find SNP variants with pleiotropic effects. The statistical formulation of factor analysis is:

$$y_{ij} = l_{i1} * f_{1j} + l_{i2} * f_{2j} + \dots + l_{ik} * f_{kj} + e_{ij},$$

where y is an observed variable i for subject j which can be expressed as a weighted composite of latent factor scores (f_{kj}) k for subject j weighted by the loading (l_{ik}) of variable i in latent factor k . The factors account for the correlations among variables in the model. In R we used the function FACTANAL and in SAS we used the procedure FACTOR, both with VARIMAX rotation to perform factor analysis. The R and SAS routines produced very similar sets of factor scores when both were applied to the same data using the same options. Loading coefficients are shown in Supplemental Table 1. For example, for the *obesity-insulin* factor of HyperGEN whites, contributions (loadings) were mainly from BMI (0.94), PBF (0.77), WC (0.86), WHR (0.40), INS (0.46) and to a lesser extent GLUC (−0.26), HDLC (−0.11) and TG (0.11). The loadings coefficients represent correlation coefficients between the “original” (original or transformed and standardized) variables and the latent factor scores. Boundaries range from −1 to +1, because loading coefficients represent correlation coefficients between factor scores and original variables, when the original variables are normally distributed and standardized $N(0,1)$. The sign can reflect the correlation direction between a factor score and original trait, but also could reflect a changed direction result of a transformation to render trait measures’ distribution to normal. In the above factor example, the negative loading of HDLC (−0.11) shows its opposite direction compared to TG (0.11), but the negative sign of GLUC (−0.26) to this factor reflects only that this trait’s measures were transformed by $1/\text{GLUC}^2$ to a normal distribution. A loading coefficient 0.4 was considered an indication of a significant contribution of a trait to a factor.

Linkage analysis was performed by multipoint variance components analysis using SEGPATH [21], detailed elsewhere [8]. Association tests between an SNP and a response variable, a latent factor score or an original trait were performed based on Linear Mixed Effects model (LME). The associations were performed using the MIXED procedure of SAS v. 9.2 for Linux OS. The LME model is given by: $Y = XB + ZU + e$, where Y is an $m \times 1$ vector of responses; X is an $m \times p$ design matrix of the fixed effects; B is the parameter $p \times 1$ vector of fixed effects; Z is an $m \times q$ incidence matrix of random effects, and U is a $q \times 1$ vector of random effects with $E(U) = \mathbf{0}$, and covariance matrix G ; e is an $m \times 1$ vector of random effects with $E(e) = \mathbf{0}$ and covariance matrix R . The fixed effects were represented by SNP genotypes (recoded as additive effects) and sex. Pedigree id was used as a random

effect to account for familial relationship. The SAS mixed model options used were $\text{ddfm} = \text{KR}$ for degrees of freedom [22] and $\text{type} = \text{un}$ for variance–covariance matrix. An SNP's additive effect was tested to determine, whether it was significantly different from zero. This was facilitated by a Bonferroni corrected threshold P value 3.98×10^{-5} equivalent to $-\log_{10} P = 4.4$, based on number of HapMap tagSNPs produced with Haplo-View in the 3p26-25 region extending from 1 to 5 Mbps, and by a $-\log_{10} P = 4.9$ considering that 12 Mbps represents the full human QTL region.

2.3. Database searches

NCBI Entrez Genes (accessed on April 2011) was used to identify genes previously reported in association with different traits. For lipids, the search was implemented with keywords “lipid” OR “cholesterol” OR “high density lipoprotein cholesterol” OR “low density lipoprotein cholesterol” OR “dyslipidemia”; for insulin, “insulin”; for glucose, “glucose”; for blood pressure, “hypertension” OR “blood pressure” OR “systolic blood pressure” OR “diastolic blood pressure”; for obesity, “obesity” OR “adiposity” OR “waist circumference” OR “percent body fat”; for type 2 diabetes, “insulin resistance” OR “glucose intolerance” OR “type 2 diabetes” OR “diabetes mellitus”; for cardiovascular disease, “cardiovascular disease” OR “heart disease”; for inflammation: “inflammation”; and for metabolic syndrome, “metabolic syndrome”. This search produced a ‘soft’ list of probable candidate genes, because it included both genes whose function has only recently been discovered and genes with extensive evidence for association with terms listed above. Additionally, we drew information on gene $\times$ gene and protein $\times$ protein interactions from the following sources: The Biomolecular Interaction Network Database (BIND) (<http://www.bind.ca>), EcoCyc (<http://www.ecocyc.org>), HIV-1 protein interactions (<http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/>) and The Human Protein Reference Database (HPRD) (<http://www.hprd.org>). The above aggregate interaction information for the 13 genes in the focused QTL region is shown as a network in the Supplemental Figure 4. The numbers 1–13 in the figure represent the positional order of these genes in the physical human map.

2.4. Human gene expression

Published information archived in the NCBI database (Accessed on July 1st, 2011) and publicly available, <http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS3715>, was used for human gene expression analysis. Expression of genes in the 3p26-25 region was examined in a larger set of Affymetrix Human Genome U95A Array platform recorded as GDS3715, reference series GSE22309 under the heading “Insulin effects on skeletal muscle in humans”. Wu et al. [23] obtained biopsies from 20 insulin sensitive, 20 insulin resistant and 15 diabetic patients before and after hyper-insulinemic–euglycemic clamps. For our purpose, online NCBI graphical presentation tools for gene expression data were utilized.

2.5. Simulation

For a better understanding of this study's findings, we simulated 100 replicates, of a trait labeled HDLC, based on the structure of 6476 subjects in the Framingham HS. The simulated data were assigned to families spanning 3 generations. We modified 10 SNPs from the Framingham HS genotypes (Affymetrix 5.0 and Affymetrix human focused gene centric chip) on chromosome 11. Nine were assigned as causative SNPs, each contributing about 2% heritability to variation in the simulated phenotype, and one was assigned as non-polymorphic. These SNPs were selected to be positioned in three nearby locations within 1 Mbps (Supplemental Figure 6). Linkage analysis was performed using Solar software [24], on a subset of the available SNPs by reducing the number of SNPs to one SNP every 1 cM. The linkage peak produced a mean LOD score of 3.24 and a standard deviation of 1.39.

2.6. Mouse population and phenotypes

The mice that generated this QTL were from the F₁₆ generation of an LG/J × SM/J Advance Intercross Line (AIL; Wustl:LG, SM-G16). Details of the AIL and of animal husbandry are published elsewhere [25]. The F₁₆ population consisted of 1002 animals partitioned and fed low-fat (247 males; 254 females) or high-fat (253 males; 248 females) diets. The diets were chosen to be isocaloric and similar in nutritional content with the exception of percent fat (15% in the low-fat diet, Harlan Teklad catalog TD88137; 43% in the high-fat diet, Research Diets catalog D12284).

The diabetes-related phenotypes mapping to this QTL are described in detail elsewhere [9]. Briefly, animals were necropsied at 20 weeks of age and serum samples were obtained via cardiac puncture and processed to measure serum glucose and insulin levels described in detail elsewhere [25]. A subset of animals was subjected to an intraperitoneal glucose tolerance test at age 20 weeks, wherein a 4-h fasting glucose level was measured followed by injection of 0.01 ml of 10% glucose solution per gram of body weight. Subsequent measurements of glucose levels taken over the course of 2 h were used to calculate the area under the curve (AUC₂₀).

2.7. Mouse genotypes and QTL mapping

DNA was extracted from liver tissue and 1536 SNPs were selected from the CTC/Oxford SNP survey and scored using the Illumina Golden Gate Assay. SNP typing was performed at the Washington University Genome Institute and 1402 autosomal SNPs were reliably scored and used for QTL mapping [9]. A genetic map in cM was generated for these SNPs based on their physical position according to the mouse genome (mm9; NCBI build37). Ordered genotypes were reconstructed at each marker using the ILP algorithm as implemented in PedPhase 2.1 [26]. Additive, dominance and imprinting genotypic scores were assigned to each marker as described elsewhere [9]. Additionally genotypes were imputed every 1 cM between the markers using Haley and Knott regression [27].

QTL analyses were performed using PROC MIXED with maximum likelihood option in SAS 9.2 (SAS Institute, Cary, NC). The full mapping model explains variation in trait (Y) using the following linear equation:

$$Y_{ijklm} = \mu + Sex_i + Diet_j + aX_{ak} + dX_{dl} + iX_{im} + sd(Sex_i \times Diet_j) \\ + as(X_{ak} \times Sex_i) + ds(X_{dl} \times Sex_i) + is(X_{im} \times Sex_i) \\ + ad(X_{ak} \times Diet_j) + dd(X_{dl} \times Diet_j) + id(X_{im} \times Diet_j) \\ + asd(X_{ak} \times Sex_i \times Diet_j) + dsd(X_{dl} \times Sex_i \times Diet_j) \\ + isd(X_{im} \times Sex_i \times Diet_j) + e_{ijklm}$$

where X_a , X_d , and X_i refer to the additive, dominance and imprinting genotype scores, respectively, μ is the population mean, and e is the residual. The direct effects of the genomic locations and their two- and three-way interactions with sex and diet were considered fixed effects. Family and its two- and three-way interactions with sex and diet were included as random effects. The $-2 \ln(\text{likelihood})$ of this model was compared to a null model including only sex, diet and sex-by-diet interaction terms using a chi-square test with 12 degrees of freedom. Probabilities were transformed into $\text{LOD} = -\log_{10}(P\text{value})$. The regression coefficients are the additive [$a = (G_{LG/LG} - (G_{SM/SM}))/2$], dominance [$d = ((G_{LG/SM} + G_{SM/LG}) - (G_{LG/LG} - G_{SM/SM}))/2$] and imprinting [$i = (G_{LG/SM} - G_{SM/LG})/2$] genotypic values. G refers to the mean phenotype of all individuals sharing the subscripted genotype, and their interactions with sex (s) and/or with diet (d). The 9 cohorts examined here are: 1) full F₁₆ population; 2) high-fat fed individuals; 3) low-fat fed individuals; 4)

males; 5) females; 6) high-fat fed males; 7) low-fat fed males; 8) high-fat fed females; 9) and low-fat fed females.

The Li and Ji method was used to calculate the number of independent tests [28]. This was then used to calculate Bonferroni adjusted significance thresholds at both genome-wide (LOD = 3.90), and chromosome-wise (LOD = 2.80 for chromosome 6) levels. Results of the entire genome-wide scan for diabetes-related traits in this population are reported elsewhere [9].

2.8. Mouse gene expression

At twenty weeks of age and after a 4-h fast animals were euthanized and necropsied. Brown fat, white fat, skeletal muscle (gastrocnemius), heart, kidney, and liver were collected from 4 males and 4 females from each strain and diet. Tissue was frozen in liquid nitrogen and stored at -80°C until extraction. RNA was extracted using RNeasy 96 Universal Tissue extraction kits (Qiagen, Valencia, CA), and quantified using a Nanodrop 2000 (Thermo Scientific, Wilmington, DE). Samples were submitted to the Washington University Microarray Core Facility, where quality was assessed using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). RNA was reverse transcribed and amplified using an Illumina Total Prep amplification kit (Ambion, Austin, TX) and then hybridized onto Illumina WG-6 v.2 BeadChips (Illumina, San Diego, CA). Arrays were scanned using the Illumina Beadstation 500, and images were processed using Illumina BeadScan software. Intensity values were analyzed using Illumina BeadStudio.

Raw data were examined using the LUMI package in R [29] and data were transformed using a variance stabilizing transformation [30]. Genes that were not expressed in each tissue were filtered out prior to analysis. Expressed genes were analyzed using Partek Genomics software v6.5 (Partek, St. Louis, MO). Significant differences in gene expression were assessed using a 3-factor ANOVA, testing for the main effects of diet, of sex, of strain and of their interactions and correcting for multiple tests using the False Discovery Rate (FDR) approach [31].

2.9. Mouse SNPs and Polyphen

Whole-genome sequencing for the LG/J ($\approx 20\times$ haploid coverage) and the SM/J ($\approx 14\times$ haploid coverage) inbred mouse strains was completed by the Washington University School of Medicine's Genome Institute using Illumina sequencing in two steps as described elsewhere [32,33]. The reference genome used was the assembly of NCBI build 37/mm9. Illumina reads from liver tissue from a single LG/J female and a single SM/J female were aligned to the reference genome using MAQ [34]. High-quality SNPs for each strain were called using SamTools [35], requiring three or more reads and an SNP quality score ≥ 20 . Polymorphisms were annotated using RefSeq coordinates downloaded from the UCSC genome browser accessed May 2010 [36]. Polyphen predictions were made by indicating option '-n mouse' when preparing the local copy of the protein database UniProtKB (www.uniprot.org) [37]. LG/J and SM/J SNPs have been submitted to dbSNP for public use under the handle 'Cheverud' [38].

3. Results

3.1. Human linkage and association results in the 3p26-25 region

Latent factor scores were used as an instrument in identifying pleiotropic effects and also in reducing the number of tests performed (Supplemental Table 1). Only after the tests were parsed to significant findings were the remaining variants retested for associations with single MetS risk traits. Consequently, results follow from latent factors to single traits.

Among the latent factors, *obesity–insulin* in HyperGEN Caucasians was most representative of obesity and to a lesser extent of insulin and glucose. While the linkage peak for this latent factor on chromosome 3p26-25 passed the LOD score threshold of 3 (Fig. 1A), linkage analysis of contributing single traits produced smaller peaks in the region with LOD scores of 1.8 for BMI, 1.2 for WC, 0.6 for PBF, and 0.25 for both INS and GLUC. Subsequent association analyses were performed by fitting an additive linear mixed effects (LME) model among SNPs and any of the 4 latent factors from each study (see Methods). Three SNPs showed significant associations, rs6790607 ($-\log_{10}P = 6.51$) and rs6788217 ($-\log_{10}P = 6.62$) both located downstream of *CNTN6* and positioned among two ribosomal protein pseudo genes (*RPL23AP38*, *RPL23AP39*), and rs17023935 ($-\log_{10}P = 6.15$) located in the intronic sequence of *CNTN4*. These SNPs had minor allele frequencies (MAF) ranging from 12% to 25% (Fig. 1C and Supplemental Table 2). Factors underlying these associations were *obesity–insulin* for Hyper-GEN Caucasians (rs17023935) and *lipids* for Family HS time 2 (rs6790607, rs6788217). Another group of significant SNP associations positioned among *IL5RA*, *TRNT1*, *CRBN*, and *LRRN1* genes and included rs3856852 ($-\log_{10}P = 4.93$), rs17031343 ($-\log_{10}P = 7.77$), rs1349311 ($-\log_{10}P = 4.88$), rs2129896 ($-\log_{10}P = 5.54$) and rs1601876 ($-\log_{10}P = 4.52$), had MAFs ranging from 15% to 24%. Factors underlying these associations were *central obesity* (rs3856852) and blood pressure (*BP*) (rs1349311, rs2129896) for Family HS time 1, for Family HS time 2 (rs1601876) and HyperGEN Caucasians (rs17031343) (Fig. 1C and Supplemental Table 2). Due to the composite nature of the multivariate latent factors, follow-up tests were performed on individual traits with these candidate variants for each study, to tease out the traits making the highest contributions to the associations (Fig. 2 and Supplemental Table 3). In the HyperGEN Caucasians, rs17023935 in an intron in *CNTN4* (3,016,284 bps, MAF 12%, *obesity–insulin* factor) had significant pleiotropic associations with BMI, WC and PBF and weaker associations with WHR and INS (Supplemental Table 3a). In Family HS time 1 this SNP showed suggestive associations with INS, HDLC and TG (Supplemental Table 3b). SNP rs6790607 intergenic between *CNTN6–CNTN4* (1,716,491 bps, MAF 20.5%, *lipids* factor), had significant pleiotropic associations with TG and HDLC in Family HS time 2 and lesser association with WC, WHR, INS, HDLC, and TG in the Family HS time 1 (Supplemental Table 3b). Finally, SNPs rs9861093 (3,385,806 bps, 164 kbps upstream of *CRBN* and 195 kbps downstream of *TRNT1*; MAF 3.5%, HyperGEN whites, *BP* factor), rs2129896 (3,649,300 bps, 192 kbps upstream of *LRRN1*; MAF 15.4%, Family HS time 1, *BP* factor), and rs17031343 (3,470,460 bps, 234 kbps upstream of *IL5RA* and 249 kbps of *CRBN*; MAF 3.4%, HyperGEN Caucasians, *BP* factor) associated significantly with systolic (SBP) and diastolic blood pressure (DBP). Furthermore, a simulation of 100 replicates was built to test the conceivable origin of a linkage peak (details provided in the Methods and Discussion sections).

3.2. LG/J × SM/J linkage findings

The mouse pleiotropic QTL shown in Fig. 1b is homologous to human chromosome 3p26-25 region and is associated with variation in three diabetes-related traits; serum INS, and GLUC levels and response to a GLUC challenge at age 20 weeks (AUC_20wks) [9]. The support interval, defined by a 1-LOD drop from the peak of the QTL, spans about 4 Mbps and contains 6 genes, *Cntn6*, *Cntn4*, *Il5ra*, *Trnt1*, *Crbn*, and *Lrrn1* (NCBI build 37, chr6: 104,433,962–108,005,522; with QTL peak at 107,015,261 bps). Because the LOD scores remained above genome-wide significance (LOD = 3.90) well beyond this interval, we examined a larger genomic region (chr6: 102,113,300–111,517,224) containing genes *Cntn6*, *Cntn4*, *Il5ra*, *Trnt1*, *Crbn*, *Lrrn1*, *Setmar*, *Sumf1*, *Iipr1*, *Bhlhe40*, *Edem1*, *Arl8b*, and *Grm7*. This larger syntenic region in the mouse remains homologous with the focal human 3p26-25 chromosomal region described above.

The genetic effects at this QTL are complex and show evidence for modulation by dietary environment and/or sex. In high fat-fed males, serum INS levels and area under the curve (AUC_20wks) show additive effects where mice homozygous for the LG allele have higher INS levels and poorer response to a GLUC challenge. This indicates that insulin is not acting efficiently to clear excess glucose from the bloodstream (i.e. insulin resistance, a precursor of type 2 diabetes and a key component of MetS). SM homozygotes have higher GLUC levels in low-fat fed females, yet when the sample is pooled there are under dominant genetic effects with no additivity, because the effects in the smaller cohort ($n = 254$ low-fat fed females versus $n = 1002$ in the full sample) are washed out (Table 1 and Supplemental Figure 3). There are also bipolar dominance imprinting effects for AUC_20wks and over dominant genetic effects for INS in mice fed a low-fat diet. Thus the mouse locus homologous with human 3p26-25 genomic region shows evidence of responsiveness to the dietary environment.

3.3. Mouse model gene expression

Gene expression analysis identified three genes with significant differential expression between LG/J and SM/J among the tissues examined (brown fat, white fat, gastrocnemius muscle, heart, kidney and liver) (Fig. 2 and Supplemental Tables 4a–f). *Crbn*, which encodes a protein related to the Lon protease family, had higher expression in LG/J in brown fat, kidney and heart tissues. *Setmar* is a fusion of a histone methylase and transposase protein and had higher expression in LG/J in brown fat and muscle tissues. *Arl8b* had higher expression in LG/J in kidney, heart and muscle tissues. This gene has been previously shown to be expressed in the mouse brain [39]. However, there are no data available on the expression differences between LG/J and SM/J for brain tissue.

4. Discussion

The multi factorial etiology of MetS involving important main effects, as well as complex epistatic and gene-by-environment interactions, reduces the likelihood that any single method will discover the multitude of underlying gene effects. Association is a statistical statement about co-occurrence of alleles and/or genotypes with phenotypes. Linkage is a statistical statement about loci showing co-segregation of disease and/or quantitative traits within families. In this study both linkage and association were used to dissect a focal genomic region associated with MetS components. While one presumes both methods to converge on true genetic effects, association is more powerful in discovering genetic subtleties than linkage. Using both methods in conjunction with gene expression, indicates that the human 3p26-25 and its homologous region on mouse chromosome 6 is associated with variation in multiple MetS components including obesity, serum lipids, insulin and glucose levels, and blood pressure. Significantly associated variants were co-localized within a gene or its neighborhood, although most of the time associations were not positioned in identical locations, which may reflect differences in sampled human cohorts. These cohorts had different MetS prevalence: 34% in African Americans and 39% in Caucasians of HyperGEN [8]; 17.1% in the Family HS time 1 and 28.8% in time 2 [15]; 23.6% in the Framingham HS Offspring cohort (exam 7) and 14.2% in the Generation 3 cohort (exam 1) [17], depending on criteria and time of recruitment.

While not all significant variants replicated across these human studies, associations of interest were identified in this focal region in at least two studies. Variants clustered in *CNTN4* and its neighborhood had significant pleiotropic effects on obesity traits. The Framingham HS has reported an intron of *CNTN4* in association with diastolic blood pressure (rs4370013, 2,654,691 bps; $P = 6 \times 10^{-6}$) [40] using 100K SNP chip data. Another region of interest was confined by *IL5RA*, *TRNT1*, *CRBN*, *LRRN1*, and included variants with significant pleiotropic effects on BP traits. Other significant associations with BP were

found for variants in the area between *EDEMI* and *GRM7* (Supplemental Tables 2 and 3b). In mouse, a non-synonymous SNP (106,688,138 bps) between LG/J and SM/J falls in *Il5ra* and changes the amino acid from arginine to cysteine (Arg⁷-to-Cys⁷). This mutation is classified as ‘possibly damaging’ by the Polyphen algorithm [37] (Supplemental Table 5a and 5b). Between the LG/J and SM/J mouse strains, *Crbn* showed differential expression in the brown fat and, together with *Arl8b*, also in kidney, heart and skeletal muscle. *Setmar* showed differential expression in brown fat and skeletal muscle (Fig. 2 and Supplemental Table 4 a–f). In the literature there is no obvious relationship among *TRTN1*, *CRBN*, *LRRN1*, *ARL8B*, *SETMAR*, and *GRM7* and MetS components. However, many of these positional candidate genes have some putative function in the neural connectivity of the brain.

For example, SNPs close to *CNTN4* have been previously associated with an index of brain structure in Alzheimer’s disease (rs10510217, 1,839,693 bps; $P = 3 \times 10^{-6}$) [41]. Further, CNVs disrupting the *CNTN6/CNTN4* region have been associated with anorexia nervosa [42]. *CNTN6/CNTN4* encode axon-associated cell adhesion molecules that function in neuronal network formation and plasticity. *IL5RA* has been associated with occurrence of ischemic stroke in individuals with T2D [43]. *Arl8b* is expressed in brain and it may be important in neural formation [39]. Mutations in *CRBN* have been associated with brain development disruption and non-syndromic mental retardation [44]. *GRM7* is a major excitatory neuro-transmitter in the central nervous system. Recent research has found a significant association between neural connectivity in the brain and glucose homeostasis and obesity [45,46]. The physiological connection between the brain and weight regulation is well known. For example, the adipokine leptin regulates satiety via receptors in the hypothalamus [47]. Other research has identified connections between neural circuitry and variation in metabolic disorders [48], and the link between Alzheimer’s disease, obesity and T2D is well studied [49,50].

Several other linkage findings have been reported for this focal region. A linkage peak was reported close to D3S1304 (6,919,306–6,919,572 bps), near *ITPR1* and *GRM7* with an LOD score of 2.5 for a cognitive dietary restraint (eating behavior) score. This linkage peak was produced from an analysis of 624 adults in 28 large families participating in the Amish Family Diabetes Study [51]. Two linkage peaks at D3S2387 (1,036,272–1,036,460 bps), close to *CNTN6*, for BMI (LOD = 3.67) and for subcutaneous adipose tissue (LOD = 3.17), were produced from an analysis of 21 African American extended pedigrees as, part of the Insulin Resistance and Atherosclerosis Family Study [52]. In the same location, the HERITAGE Family Study reported a linkage peak for abdominal subcutaneous fat (LOD = 2.16) in 215 subjects comprising 105 African American families [53]. Moreover, this focal region was interrogated for replication in an independent dataset for two traits BMI and WHR, recently made public by the GIANT consortium (accessed on July 02, 2011) http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. A handful of SNPs associated with BMI, rs2134358 (2,472,785 bps, $P = .000775$) in an intron of *CNTN4*, and rs2173488 (5,312,483 bps, $P = .000814$), located downstream of *ARL8B* [54]; and with WHR, rs7623683 (2,816,203 bps, $P = .00017$) in an intron of *CNTN4*, rs1596472 (4,163,562 bps, $P = .00047$) an intergenic variant between *LRNN1* and *SUMF1*, and rs4684484 (5,462,541 bps, $P = .00015$) located downstream of *ARL8B* [55]. In mice, two previously reported QTLs map to this region: *Hdlq1* associated with HDLC and total cholesterol in a B6 × D2 cross (23–51 cM, LOD = 2.8) and with HDLC in CAST × D2 cross (40–60 cM, LOD = 4.2) [56,57]; and *Obwq3* associated with obesity and body weight in an SM × NZB cross (peak at 42 cM, LOD = 5.8) [58]. These multiple lines of evidence show that the focal region contributes to a number of risk factors for MetS. It is worth noting that *Pparg* is located at about 52.7 cM on the mouse map [56], corresponding to about 12 Mbps on the human map, which is past the distal end of our human and mouse QTL. A large body

of evidence exists for *PPARG* involvement in obesity [59], but since this gene falls outside the focal region, it is more likely that variants between *CNTN6* and *GRM7* are contributing to the associations we identify.

This focused analysis of 3p26-25 revealed several variants significantly associated with obesity, lipids and/or blood pressure. Associations with insulin and/or glucose were nominal. However, given the strength of the homologous mouse insulin and glucose QTL, and the fact that disordered glucose metabolism is an important component of MetS [60], it was of interest to interrogate the available literature and interaction databases to determine if any genes in this focal region had prior evidence of interacting with insulin and glucose in human studies (see Methods). Two genes, *ITPR1* and *BHLHE40*, had prior evidence of interaction with insulin. Supplemental Figure 4 depicts a gene network that includes these two genes. *ITPR1* (inositol 1,4,5-trisphosphate receptor, type 1) is classified as a member of “Regulation of insulin secretion” pathway in NCBI BioSystems (source: REACTOME [REACT_18325]) and interacts with *RYR2* (Supplemental Figure 4), which is a candidate for heart disease and blood pressure control [61]. One way insulin exerts its multiple biological actions is through gene expression regulation. *BHLHE40* (basic helix–loop–helix family, member e40) is a transcription factor that shows a noticeable gene expression pattern in the GDS3715 record (the NCBI Gene Expression Omnibus database) in biopsies from human skeletal muscle (vastus lateralis) of 20 insulin sensitive, 20 insulin resistant and 15 diabetic patients before and after a hyperinsulinemic–euglycemic clamp study [23]. Among the insulin-sensitive samples, the authors identified 779 insulin responsive genes, including 70 transcription factors. They distinguished 4 genes, *RRAD*, *IGFBP5*, *INSIG1*, and *NR4A1*, as significantly up-regulated by insulin in cultured muscle cells. Supplemental Figure 5 shows expression patterns of *CNTN6*, *ITPR1*, and *BHLHE40* compared to *NR4A1*, one of the top insulin up-regulated genes resulting from the clamp study [23]. *BHLHE40* had a similar up-regulated pattern to *NR4A1* in these patients. In contrast, *ITPR1* was relatively down-regulated, while *CNTN6* expression was not influenced by insulin treatment. Recently, the *BHLHE40* has been implicated as an important intermediary of insulin action on transcription in human muscle [62] and is involved in circadian rhythm and metabolism [63]. These results suggest that candidate genes in 3p26-25 are associated with insulin, despite the stronger association of SNPs in this region with the obesity, lipids and/or blood pressure traits in the human data examined.

The nominal association with insulin in the human data may be a function of the power to detect associations if genes interact in specific sex and/or dietary environments. The mouse results presented here are highly context dependent, showing that gene-by-environment interactions are an important feature of MetS components. Indeed, gene-by-environment interactions are not always consistent among genotypes, across environments, or even among traits. Supplemental Figure 3 illustrates that the genetic effects at this focal QTL vary among the different sex-by-diet cohorts in the mouse experiment. If two cohorts have opposite genetic effects (for example, if in one cohort the SS genotype results in higher insulin levels and in a second cohort the same SS genotype results in lower insulin levels) and they are pooled together in an analysis, the effects can wash each other out and go undetected. This has important implications for human studies where it is difficult, if not impossible, to stratify a sample into sex-by-diet cohorts and maintain sufficient statistical power to detect associations [64]. Thus using mouse results to leverage associations in human studies is a way to identify candidates that might otherwise remain undetected due to statistical power constraints. This is particularly true for complex multifactorial components of MetS.

Our study’s strength originates from the hypothesis-driven approach in these analyses. It identified associations with small effects contributing to MetS components that would have

not been deemed significant if this study had been conducted at the genome-wide level. To determine whether these findings can be expected theoretically, we simulated 100 replications of a QTL constructed to mimic 9 common (MAF > 40%) polymorphic causative SNPs placed within a 1 Mbps region, each with a heritability of 0.02 (see Methods). The simulated QTL had a mean LOD score linkage peak of 3.24 and a standard deviation of 1.39, supporting our analysis and indicating that the original HyperGEN linkage peak for obesity–insulin factor can be produced by several variants contributing small effects (Supplemental Figure 6). While this theoretical model supports the ‘common disease–common allele’ findings for the 3p26-25 region, our study also has a weakness, that it was not designed to analyze rare variants. Rare variants are expected to contribute significantly to the modulation of complex metabolic traits [65,66].

A number of candidate genes in 3p26-25 reported in our study have functions in the brain. While the precise role of these genetic variants on brain function in relation to MetS components remains to be established, the brain’s role in affecting MetS components has clinical significance, particularly when considering gene-by-environment interactions. It is documented that short-term exercise intervention can induce favorable changes in body composition, but biologically significant changes if sustained for longer periods [67–69]. Yet while promoting healthy behavior (exercise and healthy diet) in children and adults is important [70], dieting as a behavioral change to control weight has a high failure rate (>80%) [71] in the sense of regaining weight. These reverse effects (weight regain) were recently attributed to the stress associated with dieting. In mice these reverse effects were physiologically associated with changes in a brain area where stress and reward centers intersect. Weight regain was associated with less activity and higher stress-induced corticosterone levels and only previously dieting mice showed significant increases in orexigenic hormones, melanin-concentrating hormone, and orexin levels in response to high fat food [72]. Thus, based on this perspective our study findings warrant a sequel with sequencing to address rare allele effects in this region. Moreover, functional assays on these genes will help to better understand the relationship between MetS and brain. Epigenetics, testing effects of diet and development of reliable biomarkers for MetS and its components should provide a better understanding of the 3p26-25 region’s contributions to MetS. Our approach of incorporating multiple lines of evidence highlighted new and potentially clinically relevant associations that may have implications for personal genetic-based prevention and treatment strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported in part by several grants from the National Heart and Blood Institutes (NHLBI): the HyperGEN Study grant U01 HL54471, the Family Heart Study (grant HL0877700), Genetic Determinants of LVH Phenotype (grants R01 HL071782, R21HL094668), a post-doctoral training grant in Genetic Epidemiology (grant T32-HL091823); grants from the National Institute of Diabetes and Digestive Kidney Disease (NIDDK) (grant R01 DK055736), by Washington University Diabetes Research and Training Center (grant DK056341), and by NIH (grant P30 DK056341).

The authors express their gratitude to The Framingham Heart Study, the NHLBI, and Boston University for providing access to the Framingham HS data through the NCBI’s dbGaP. The authors thank Drs. Warwick Daw and Yun Ju Sung for their active participation and discussions in setting the simulation presented in this article.

Abbreviations

MetS	metabolic syndrome
T2D	type 2 diabetes mellitus
BMI	body mass index
WC	waist circumference
WHR	waist-to-hip ratio
PBF	percent body fat
INS	fasting insulin
GLUC	fasting glucose
TG	fasting triglycerides
HDL	high density lipoprotein cholesterol
LDL	low density lipoprotein cholesterol
SBP	systolic blood pressure
DBP	diastolic blood pressure
BP	blood pressure
QTL	quantitative trait locus
LOD	logarithm of base 10 of odds for linkage evidence
GWAS	genome-wide association scans

References

1. Mozumdar A, Liguori G. Persistent increase of prevalence of metabolic syndrome among U.S. adults: NHANES III to NHANES 1999–2006. *Diabetes Care*. 2011; 34:216–9. [PubMed: 20889854]
2. Sullivan PW, Ghushchyan V, Wyatt HR, et al. Productivity costs associated with cardiometabolic risk factor clusters in the United States. *Value Health*. 2007; 10:443–50. [PubMed: 17970926]
3. Monda KL, North KE, Hunt SC, et al. The genetics of obesity and the metabolic syndrome. *Endocr Metab Immune Disord Drug Targets*. 2010; 10:86–108. [PubMed: 20406164]
4. Djousse L, Padilla H, Nelson TL, et al. Diet and metabolic syndrome. *Endocr Metab Immune Disord Drug Targets*. 2010; 10:124–37. [PubMed: 20088815]
5. Loos RJ, Katzmarzyk PT, Rao DC, et al. Genome-wide linkage scan for the metabolic syndrome in the HERITAGE Family Study. *J Clin Endocrinol Metab*. 2003; 88:5935–43. [PubMed: 14671193]
6. Clerget-Darpoux F, Elston RC. Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. *Hum Hered*. 2007; 64:91–6. [PubMed: 17476108]
7. Kitsios GD, Zintzaras E. Genomic convergence of genome-wide investigations for complex traits. *Ann Hum Genet*. 2009; 73:514–9. [PubMed: 19604225]
8. Kraja AT, Hunt SC, Pankow JS, et al. Quantitative trait loci for metabolic syndrome in the Hypertension Genetic Epidemiology Network study. *Obes Res*. 2005; 13:1885–90. [PubMed: 16339118]
9. Lawson HA, Lee A, Fawcett GL, et al. The importance of context to the genetic architecture of diabetes-related traits is revealed in a genome-wide scan of a LG/J × SM/J murine model. *Mamm Genome*. 2011; 22:197–208. [PubMed: 21210123]
10. Lawson HA, Cheverud JM. Metabolic syndrome components in murine models. *Endocr Metab Immune Disord Drug Targets*. 2010; 10:25–40. [PubMed: 20088816]

11. Leduc MS, Lyons M, Darvishi K, et al. The mouse QTL map helps interpret human genome-wide association studies for HDL cholesterol. *J Lipid Res.* 2011; 52:1139–49. [PubMed: 21444760]
12. Fawcett GL, Roseman CC, Jarvis JP, et al. Genetic architecture of adiposity and organ weight using combined generation QTL analysis. *Obesity (Silver Spring).* 2008; 16:1861–8. [PubMed: 18551125]
13. Multi-center genetic study of hypertension: the Family Blood Pressure Program (FBPP). *Hypertension.* 2002; 39:3–9. [PubMed: 11799070]
14. Higgins M, Province M, Heiss G, et al. NHLBI Family Heart Study: objectives and design. *Am J Epidemiol.* 1996; 143:1219–28. [PubMed: 8651220]
15. Kraja AT, Borecki IB, North K, et al. Longitudinal and age trends of metabolic syndrome and its risk factors: the Family Heart Study. *Nutr Metab (Lond).* 2006; 3:41. [PubMed: 17147796]
16. Cupples LA, Heard-Costa N, Lee M, Atwood LD. Genetics analysis workshop 16 problem 2: the Framingham Heart Study data. *BMC Proc.* 2009; 3(Suppl 7):S3. [PubMed: 20018020]
17. Park YM, Province MA, Gao X, et al. Longitudinal trends in the association of metabolic syndrome with 550 k single-nucleotide polymorphisms in the Framingham Heart Study. *BMC Proc.* 2009; 3(Suppl 7):S116. [PubMed: 20017981]
18. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21:263–5. [PubMed: 15297300]
19. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006; e190:2.
20. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010; 34:816–34. [PubMed: 21058334]
21. Province MA, Rice TK, Borecki IB, et al. Multivariate and multilocus variance components method, based on structural relationships to assess quantitative trait linkage via SEGPATH. *Genet Epidemiol.* 2003; 24:128–38. [PubMed: 12548674]
22. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics.* 1997; 53:983–97. [PubMed: 9333350]
23. Wu X, Wang J, Cui X, et al. The effect of insulin on expression of genes and biochemical pathways in human skeletal muscle. *Endocrine.* 2007; 31:5–17. [PubMed: 17709892]
24. Almasy L, Blangero J. Contemporary model-free methods for linkage analysis. *Adv Genet.* 2008; 60:175–93. [PubMed: 18358321]
25. Ehrlich TH, Kenney-Hunt JP, Pletscher LS, Cheverud JM. Genetic variation and correlation of dietary response in an advanced intercross mouse line produced from two divergent growth lines. *Genet Res.* 2005; 85:211–22. [PubMed: 16174340]
26. Li J, Jiang T. Computing the minimum recombinant haplotype configuration from incomplete genotype data on a pedigree by integer linear programming. *J Comput Biol.* 2005; 12:719–39. [PubMed: 16108713]
27. Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity.* 1992; 69:315–24. [PubMed: 16718932]
28. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity.* 2005; 95:221–7. [PubMed: 16077740]
29. Du P, Kibbe WA, Lin SM. Lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008; 24:1547–8. [PubMed: 18467348]
30. Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.* 2008; e11:36.
31. Storey JD. A direct approach to false discovery rates. *J R Stat Soc Ser B.* 2002; 64:479–98.
32. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med.* 2009; 361:1058–66. [PubMed: 19657110]
33. Ding L, Ellis MJ, Li S, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature.* 2010; 464:999–1005. [PubMed: 20393555]
34. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 2008; 18:1851–8. [PubMed: 18714091]

35. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAM tools. *Bioinformatics*. 2009; 25:2078–9. [PubMed: 19505943]
36. Rhead B, Karolchik D, Kuhn RM, et al. The UCSC Genome Browser database: update 2010. *Nucleic Acids Res*. 2010; 38:D613–9. [PubMed: 19906737]
37. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002; 30:3894–900. [PubMed: 12202775]
38. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001; 29:308–11. [PubMed: 11125122]
39. Haraguchi T, Yanaka N, Nogusa Y, et al. Expression of ADP-ribosylation factor-like protein 8B mRNA in the brain is down-regulated in mice fed a high-fat diet. *Biosci Biotechnol Biochem*. 2006; 70:1798–802. [PubMed: 16861820]
40. Levy D, Larson MG, Benjamin EJ, et al. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC Med Genet*. 2007; 8(Suppl 1):S3. [PubMed: 17903302]
41. Furney SJ, Simmons A, Breen G, et al. Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. *Mol Psychiatry*. 2010
42. Wang K, Zhang H, Bloss CS, et al. A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. *Mol Psychiatry*. 2010
43. Luk AO, Wang Y, Ma RC, et al. Predictive role of polymorphisms in interleukin-5 receptor alpha-subunit, lipoprotein lipase, integrin A2 and nitric oxide synthase genes on ischemic stroke in type 2 diabetes—an 8-year prospective cohort analysis of 1327 Chinese patients. *Atherosclerosis*. 2010
44. Higgins JJ, Hao J, Kosofsky BE, Rajadhyaksha AM. Dysregulation of large-conductance Ca²⁺-activated K⁺ channel expression in nonsyndromal mental retardation due to a cereblon p. R419× mutation. *Neurogenetics*. 2008; 9:219–23. [PubMed: 18414909]
45. Schwartz MW, Porte D Jr. Diabetes, obesity, and the brain. *Science*. 2005; 307:375–9. [PubMed: 15662002]
46. Thorens B. Central control of glucose homeostasis: the brain–endocrine pancreas axis. *Diabetes Metab*. 2010; 36(Suppl 3):S45–9. [PubMed: 21211735]
47. Elmquist JK, Elias CF, Saper CB. From lesions to leptin: hypothalamic control of food intake and body weight. *Neuron*. 1999; 22:221–32. [PubMed: 10069329]
48. Levin BE. The obesity epidemic: metabolic imprinting on genetically susceptible neural circuits. *Obes Res*. 2000; 8:342–7. [PubMed: 10933311]
49. Profenno LA, Porsteinsson AP, Faraone SV. Meta-analysis of Alzheimer's disease risk with obesity, diabetes, and related disorders. *Biol Psychiatry*. 2010; 67:505–12. [PubMed: 19358976]
50. Naderali EK, Ratcliffe SH, Dale MC. Obesity and Alzheimer's disease: a link between body weight and cognitive function in old age. *Am J Alzheimers Dis Other Dement*. 2009; 24:445–9. [PubMed: 19801534]
51. Steinle NI, Hsueh WC, Snitker S, et al. Eating behavior in the Old Order Amish: heritability analysis and a genome-wide linkage analysis. *Am J Clin Nutr*. 2002; 75:1098–106. [PubMed: 12036819]
52. Norris JM, Langefeld CD, Scherzinger AL, et al. Quantitative trait loci for abdominal fat and BMI in Hispanic–Americans and African–Americans: the IRAS Family study. *Int J Obes (Lond)*. 2005; 29:67–77. [PubMed: 15534617]
53. Rice T, Chagnon YC, Perusse L, et al. A genomewide linkage scan for abdominal subcutaneous and visceral fat in black and white families: the HERITAGE Family Study. *Diabetes*. 2002; 51:848–55. [PubMed: 11872690]
54. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010; 42:937–48. [PubMed: 20935630]
55. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature genetics*. 2010; 42:949–60. [PubMed: 20935629]
56. Lyons MA, Wittenburg H, Li R, et al. Quantitative trait loci that determine lipoprotein cholesterol levels in an intercross of 129S1/SvImJ and CAST/Ei inbred mice. *Physiol Genomics*. 2004; 17:60–8. [PubMed: 14701919]

57. Su Z, Ishimori N, Chen Y, et al. Four additional mouse crosses improve the lipid QTL landscape and identify *Lipg* as a QTL gene. *J Lipid Res.* 2009; 50:2083–94. [PubMed: 19436067]
58. Stylianou IM, Korstanje R, Li R, et al. Quantitative trait locus analysis for obesity reveals multiple networks of interacting loci. *Mamm Genome.* 2006; 17:22–36. [PubMed: 16416088]
59. Rankinen T, Zuberi A, Chagnon YC, et al. The human obesity gene map: the 2005 update. *Obesity (Silver Spring).* 2006; 14:529–644. [PubMed: 16741264]
60. Farag YM, Gaballa MR. Diabetes: an overview of a rising epidemic. *Nephrol Dial Transplant.* 2011; 26:28–35. [PubMed: 21045078]
61. Kraja AT, Hunt SC, Rao DC, et al. Genetics of hypertension and cardiovascular disease and their interconnected pathways: lessons from large studies. *Curr Hypertens Rep.* 2011; 13:46–54. [PubMed: 21128019]
62. Rome S, Meugnier E, Lecomte V, et al. Microarray analysis of genes with impaired insulin regulation in the skeletal muscle of type 2 diabetic patients indicates the involvement of basic helix–loop–helix domain-containing, class B, 2 protein (BHLHB2). *Diabetologia.* 2009; 52:1899–912. [PubMed: 19590847]
63. Noshiro M, Usui E, Kawamoto T, et al. Liver X receptors (LXRalpha and LXRbeta) are potent regulators for hepatic *Dec1* expression. *Genes Cells.* 2009; 14:29–40. [PubMed: 19032342]
64. Lawson HA, Cady JE, Partridge C, et al. Genetic effects at pleiotropic loci are context-dependent with consequences for the maintenance of genetic variation in populations. *PLoS Genet.* 2011; 7:e1002256. [PubMed: 21931559]
65. Brunham LR, Singaraja RR, Hayden MR. Variations on a gene: rare and common variants in *ABCA1* and their impact on HDL cholesterol levels and atherosclerosis. *Annu Rev Nutr.* 2006; 26:105–29. [PubMed: 16704350]
66. Romeo S, Yin W, Kozlitina J, et al. Rare loss-of-function mutations in *ANGPTL* family members contribute to plasma triglyceride levels in humans. *J Clin Invest.* 2009; 119:70–9. [PubMed: 19075393]
67. Wilmore JH, Despres JP, Stanforth PR, et al. Alterations in body weight and composition consequent to 20 wk of endurance training: the HERITAGE Family Study. *Am J Clin Nutr.* 1999; 70:346–52. [PubMed: 10479196]
68. Wilmore JH, Stanforth PR, Hudspeth LA, et al. Alterations in resting metabolic rate as a consequence of 20 wk of endurance training: the HERITAGE Family Study. *Am J Clin Nutr.* 1998; 68:66–71. [PubMed: 9665098]
69. Morton AR, Stanforth PR, Freund BJ, et al. Alterations in plasma lipids consequent to endurance training and beta-blockade. *Med Sci Sports Exerc.* 1989; 21:288–92. [PubMed: 2567483]
70. Hart KH, Herriot A, Bishop JA, Truby H. Promoting healthy diet and exercise patterns amongst primary school children: a qualitative investigation of parental perspectives. *J Hum Nutr Diet.* 2003; 16:89–96. [PubMed: 12662367]
71. Wing RR, Phelan S. Long-term weight loss maintenance. *Am J Clin Nutr.* 2005; 82:222S–5S. [PubMed: 16002825]
72. Pankevich DE, Teegarden SL, Hedin AD, et al. Caloric restriction experience reprograms stress and orexigenic pathways and promotes binge eating. *J Neurosci.* 2010; 30:16399–407. [PubMed: 21123586]

Appendix A. Supplementary data

Supplementary data to this article can be found online at [doi:10.1016/j.metabol.2012.01.008](https://doi.org/10.1016/j.metabol.2012.01.008).

Fig. 1.

A. Linkage results on chromosome 3 of the HyperGEN study. B. Linkage results on chromosome 6 of generation 16 of *LG/J* × *SM/J* mouse syntenic with humans. C. Association test results from 3 human studies (HyperGEN, Framingham Heart Study and Family Heart Study) in the region 3p26-25 (for more details on Fig. 1C, see Supplemental Figures 1 and 2, and Supplemental Table 2).

Fig. 2. Linkage, association, and gene expression results in human samples and mouse. This schematic summarizes results from Supplemental Tables 2, 3a–c, and 4a–f.

Table 1

Additive, dominance and imprinting scores for traits mapping in mice QTL locus named *Ddiab6d* [9].

Trait	LOD	Cohort	Additive ^a	S.E.	P value	Dominance	S.E.	P value	Imprinting	S.E.	P value
Insulin	8.85	All ^b	-0.26	0.12	3.32E-02	-0.14	0.16	3.99E-01	0.01	0.13	9.47E-01
		Low	0.00	0.11	9.65E-01	0.31	0.16	5.59E-02	-0.08	0.12	5.09E-01
		High	-0.61	0.22	5.50E-03	-0.55	0.29	5.40E-02	0.08	0.21	7.05E-01
		Female	0.10	0.14	4.46E-01	0.18	0.18	3.28E-01	0.05	0.14	7.02E-01
		Male	-0.60	0.19	1.60E-03	-0.43	0.27	1.10E-01	-0.03	0.20	8.81E-01
		LFF	0.12	0.14	4.05E-01	0.04	0.21	8.38E-01	-0.14	0.15	3.50E-01
		HFF	0.05	0.23	8.31E-01	0.32	0.30	2.94E-01	0.24	0.21	2.50E-01
AUC_20wks	3.43	LFM	-0.12	0.16	4.76E-01	0.53	0.25	3.19E-02	-0.02	0.19	9.15E-01
		HFM	-1.19	0.34	7.00E-04	-1.38	0.47	4.20E-03	-0.10	0.34	7.64E-01
		All	-949.09	326.57	4.10E-03	-110.67	457.05	8.09E-01	-162.00	325.75	6.20E-01
		Low	-26.6614	284.80	9.26E-01	387.73	399.35	3.34E-01	-684.63	293.82	2.19E-02
		High	-1960.36	594.30	1.40E-03	-497.16	811.52	5.42E-01	362.09	562.46	5.21E-01
		Female	-346.28	239.02	1.51E-01	299.32	329.22	3.66E-01	-8.79	230.92	9.70E-01
		Male	-1503.52	589.19	1.23E-02	-574.94	845.07	4.98E-01	-229.00	613.25	7.10E-01
Glucose	3.19	LFF	-137.46	277.80	6.23E-01	375.15	410.04	3.65E-01	-408.90	300.42	1.80E-01
		HFF	-537.06	389.35	1.75E-01	246.88	510.86	6.31E-01	332.07	336.38	3.29E-01
		LFM	270.39	493.38	5.87E-01	346.13	685.49	6.16E-01	-855.65	504.2	9.69E-02
		HFM	-3077.13	1018.62	3.90E-03	-1468.30	1460.91	3.20E-01	613.95	1058.98	5.65E-01
		All	1.84	3.13	5.55E-01	-12.66	3.97	1.50E-03	1.31	3.20	6.83E-01
		Low	8.08	4.05	4.65E-02	-13.66	5.55	1.43E-02	-2.35	4.47	5.99E-01
		High	-5.60	4.32	1.96E-01	-10.67	5.63	5.87E-02	3.09	4.21	4.64E-01
		Female	6.20	4.53	1.72E-01	-19.12	5.85	1.20E-03	6.75	4.46	1.31E-01
		Male	-0.21	3.97	9.58E-01	-5.87	5.25	2.65E-01	-5.08	4.25	2.33E-01
		LFF	19.87	5.60	5.00E-04	-18.72	8.02	2.08E-02	6.10	6.00	3.11E-01
		HFF	-9.62	6.43	1.37E-01	-11.56	8.21	1.61E-01	2.07	5.94	7.28E-01
		LFM	0.36	5.05	9.44E-01	-5.93	7.19	4.11E-01	-11.29	5.83	5.45E-02
		HFM	-2.33	5.65	6.81E-01	-8.61	7.65	2.62E-01	3.72	5.79	5.22E-01

^aAdditive/Dominance/Imprinting, additive/dominance/imprinting effect; S.E., standard error of the Additive/Dominance/Imprinting effects.

^bAll, all data; Low/High, low/high fat diet; LFF/HFF, low/high fat diet in females; LFM/HFM, low/high fat diet in males.