nature genetics

Genetic variation in selenoprotein S influences inflammatory response

Joanne E Curran[1,6], Jeremy B M Jowett[2,3,6], Kate S Elliott[2,6], Yuan Gao[4], Kristi Gluschenko[2], Jianmin Wang[2], Dalia M Abel Azim[5], Guowen Cai[1], Michael C Mahaney[1], Anthony G Comuzzie[1], Thomas D Dyer[1], Ken R Walder[3,4], Paul Zimmet[2,3], Jean W MacCluer[1], Greg R Collier[3,4], Ahmed H Kissebah[5,6] & John Blangero[1,3,6]

Chronic inflammation has a pathological role in many common diseases and is influenced by both genetic and environmental factors. Here we assess the role of genetic variation in selenoprotein S (*SEPS1*, also called *SELS* or *SELENOS*), a gene involved in stress response in the endoplasmic reticulum and inflammation control. After resequencing *SEPS1*, we genotyped 13 SNPs in 522 individuals from 92 families. As inflammation biomarkers, we measured plasma levels of IL-6, IL-1β and TNF-α. Bayesian quantitative trait nucleotide analysis identified associations between *SEPS1* polymorphisms and all three proinflammatory cytokines. One promoter variant, −105G→A, showed strong evidence for an association with each cytokine (multivariate $P = 0.0000002$). Functional analysis of this polymorphism showed that the A variant significantly impaired *SEPS1* expression after exposure to endoplasmic reticulum stress agents ($P = 0.00006$). Furthermore, suppression of *SEPS1* by short interfering RNA in macrophage cells increased the release of IL-6 and TNF-α. To investigate further the significance of the observed associations, we genotyped −105G→A in 419 Mexican American individuals from 23 families for replication. This analysis confirmed a significant association with both TNF-α ($P = 0.0049$) and IL-1β ($P = 0.0101$). These results provide a direct mechanistic link between *SEPS1* and the production of inflammatory cytokines and suggest that *SEPS1* has a role in mediating inflammation.

Activation of the immune system is involved in the pathogenesis of many common complex diseases, such as cardiovascular disease, diabetes and cancer. Ongoing inflammatory insults might contribute to the development of such diseases. This inflammatory activation, a result of cellular exposure to stress conditions, is reflected by increases in levels of circulating proinflammatory cytokines. Genetic and environmental factors are likely to influence inflammatory response, but relatively little is known about the identity of the genes underlying its regulation.

We recently identified a protein, TANIS, that putatively functions in the stress responses of the endoplasmic reticulum (ER) that are linked to the immune and inflammatory signaling pathways[1,2]. The human homolog of TANIS was later identified as mammalian selenoprotein S (encoded by *SEPS1*, also called *SELS* and *VIMP*)[3,4]. Expression of *SEPS1* was dysregulated in diabetic and glucose-intolerant *Psammomys obesus*, a polygenic animal model of the metabolic syndrome[1]. Yeast two-hybrid screening identified serum amyloid A, an acute-phase inflammatory response protein, as an interacting protein of SEPS1, which was confirmed by Biacore experiments[1]. These findings were later confirmed in a study of human adipose tissue from diabetic and normal individuals after *in vivo* insulin stimulation[5]. These results suggested that SEPS1 has a role in the inflammation pathway.

The crucial role of SEPS1 in protecting the functional integrity of the ER against potential metabolic stressors was recently clarified[4]. SEPS1 has been classified as a new ER membrane protein that participates in the processing and removal of misfolded proteins from the ER to the cytosol, where they are polyubiquitinated and degraded through the proteasome. SEPS1 also regulates cellular redox balance and protects the ER against the deleterious effects of oxidative stress. A functionally impaired ER responds to these adverse events by inducing expression of a number of genes, which leads to activation of the transcription factor NF-κB. Activated NF-κB then translocates to the nucleus where it activates the transcription of genes including those that encode the proinflammatory cytokines[6].

The human gene *SEPS1* is located on chromosome 15q26.3, consists of six exons and encodes a 189–amino acid protein. This region of chromosome 15 was previously suggested to contain quantitative trait loci that influence inflammatory disorders including insulin-dependent diabetes mellitus, Alzheimer disease and celiac disease[7–10]. Therefore, *SEPS1* is both a strong functional candidate and a

[1]Southwest Foundation for Biomedical Research, San Antonio, Texas 78227, USA. [2]International Diabetes Institute, Caulfield, Victoria, Australia. [3]ChemGenex Pharmaceuticals, Geelong, Victoria, Australia. [4]Metabolic Research Unit, School of Health Sciences, Deakin University, Waurn Ponds, Victoria, Australia. [5]Take Off Pounds Sensibly Center for Obesity and Metabolic Research, Medical College of Wisconsin, Wisconsin, USA. [6]These authors contributed equally to this work. Correspondence should be addressed to J.B. (john@darwin.sfbr.org).

Received 14 June; accepted 3 August; published online 9 October 2005; doi:10.1038/ng1655

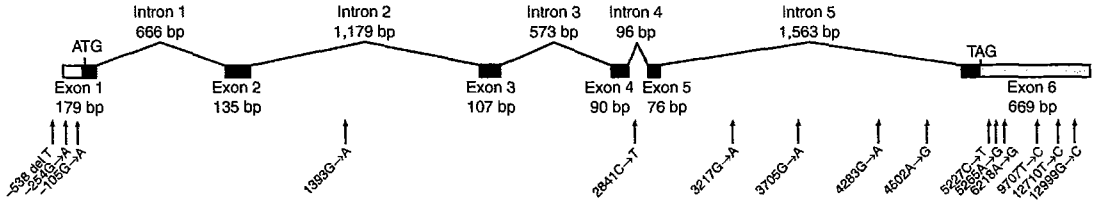

Figure 1 *SEPS1* gene structure.

positional candidate for involvement in various inflammation-related disorders.

To investigate the role of *SEPS1* in inflammation, we identified all its naturally occurring genetic variants and assessed their influence on inflammation using several phenotypic measures. Specifically, we examined the effect of *SEPS1* sequence variation on circulating levels of primary (IL-1β and TNF-α) and secondary (IL-6) proinflammatory cytokines as an index of inflammation[11].

RESULTS

Variant identification and SNP selection

Figure 1 shows the gene structure of *SEPS1*. We sequenced 9.3 kbp encompassing the putative promoter, all exons, introns and additional conserved regions, identified by comparative genomics, in 50 individuals from multiple ethnic backgrounds to increase our chances of detecting new genetic variants. We identified 24 variants, 19 of which were new (not listed in any public databases at time of publication). Only 11 of the 24 variants identified were observed in the subset of individuals from Wisconsin (**Supplementary Table 1** online). Given the sample size for the resequencing of this subset, the chance of detecting a rare population-specific variant (allele frequency of 0.01 or less) present in the population was 0.18. Therefore, we might have missed rare variants specific to this population. We also identified five additional SNPs available in the public databases in the extreme 3′ untranslated region of *SEPS1*, outside our sequencing region (**Supplementary Table 1**). We then genotyped these 16 polymorphisms (11 identified by resequencing and 5 from the public databases) in a sample of 522 individuals from 92 families from Wisconsin. The structure of the available pairwise relationships represented by this data set is given in **Supplementary Table 2** online. The frequency of the minor alleles varied from less than 0.01 to 0.32 (maximum likelihood estimates of the allele frequencies allowing for the relatedness of pedigree members are given in **Supplementary Table 1**).

Linkage disequilibrium

Three SNPs (2841C→T, 12999G→C and 13171C→G) were omitted from further analysis because fewer than four copies of the minor allele were available in the data set, precluding accurate statistical analysis. The other 13 *SEPS1* SNPs were in varying degrees of disequilibrium with each other. **Figure 2** shows the overall pattern of linkage disequilibrium as measured by the absolute correlation (ρ) among genotypes and graphed using the computer program SOLAR. This analysis included all genotyped individuals, which is appropriate when trying to assess redundant statistical information used in association analysis. Overall, linkage disequilibrium across *SEPS1* is rather modest. Based on the evaluation of the eigenstructure of the correlation matrix among markers[12], the effective number of polymorphisms is 9.7. Therefore, these 13 markers behave statistically like 10 effectively independent markers. The average pairwise correlation among SNPs was 0.25. The average multiple correlation between each marker and all other markers taken jointly was 0.74, with SNP-specific values ranging from 0.07 to 1. Conditional on the family information, we estimated the haplotypes of each individual using the computer program MERLIN[13]. We observed 55 unique haplotypes of the theoretical maximum of $2^{13} = 8{,}192$ and the sample maximum of 448 (which is equal to the number of founder genomes in the families). Twelve haplotypes had a frequency of 0.01 or greater, and the seven most frequent haplotypes were required to capture 80% of all individuals. The estimated haplotype diversity was 0.872.

Heritabilities and correlations of inflammatory markers

We carried out multivariate variance component analysis using SOLAR[14] to estimate the heritabilities and the genetic correlations among the three plasma markers of inflammation. The analysis statistically controlled for the effects of age and sex and their interactions. All three cytokines had highly significant heritabilities: 0.492 ± 0.077 ($P = 8.3 \times 10^{-11}$) for IL-1β; 0.368 ± 0.076 ($P = 6.4 \times 10^{-7}$) for IL-6; and 0.443 ± 0.079 ($P = 1.0 \times 10^{-8}$) for TNF-α. These results indicate that plasma levels of these proinflammatory cytokines have a substantial genetic component. To investigate further the pleiotropic relationship among these three phenotypes, we estimated their genetic

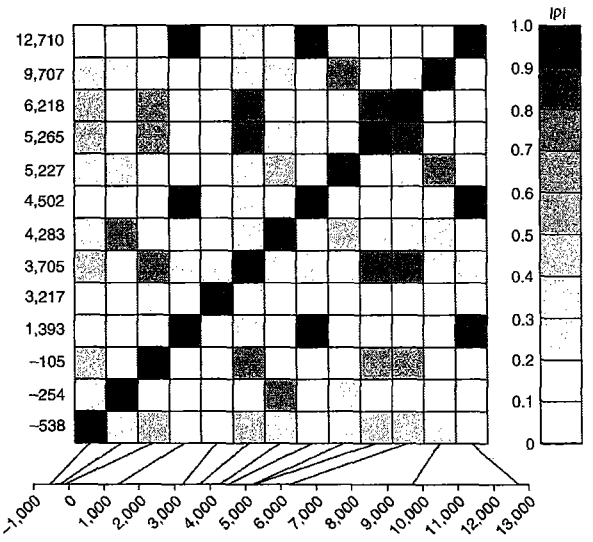

Figure 2 Pattern of linkage disequilibrium in *SEPS1*. The intensity of red color in a block indicates the magnitude of the absolute correlation (ρ) between SNP alleles (or genotypes).

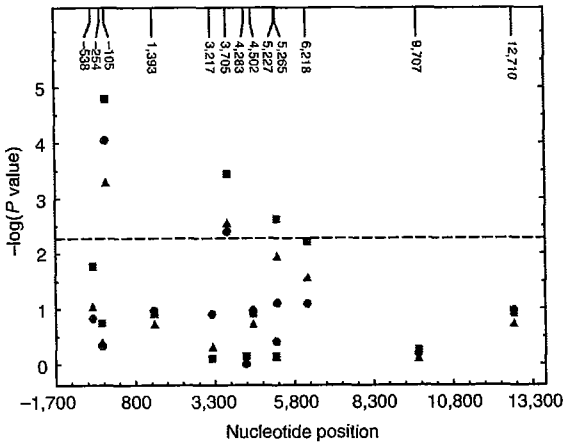

Figure 3 Results of robust measured genotype analysis for marginal associations between *SEPS1* SNPs and plasma cytokine measures. Squares, IL-1β; circles, TNF-α; triangles, IL-6. Results above the black dashed line represent those that achieved experiment-wide significance ($P < 0.05$) after controlling for multiple tests (obtained as the number of effectively independent genetic markers).

correlations with one another. The observed genetic correlations were all high: 0.898 ± 0.036 between IL-1β and IL-6; 0.958 ± 0.016 between IL-1β and TNF-α; and 0.881 ± 0.039 between IL-6 and TNF-α. Such high genetic correlations suggest the presence of genetic variation that jointly influences all these proinflammatory markers, although shared environmental effects may also have a role. These quantitative genetic analyses show that circulating levels of IL-1β, IL-6 and TNF-α are influenced by genes and that most of these genes seem to influence all three phenotypes pleiotropically.

Association of *SEPS1* genetic variants with inflammation markers

We carried out marginal association analysis using a measured genotype approach as implemented in the computer program SOLAR. First, we tested for hidden population stratification using the QTDT approach[15] in SOLAR. None of the SNPs showed statistical evidence of stratification, allowing us to use the more powerful measured genotype procedure to assess association. **Figure 3** shows the results of these association analyses (the -log(P value) for each test). Three polymorphisms (−105G→A, 3705G→A and 5227C→T) showed significant associations with at least one cytokine after correcting for multiple tests. Two SNPs, −105G→A and 3705G→A, showed significant correlations with all three proinflammatory cytokines. In particular, the −105G→A polymorphism, located in the promoter region of *SEPS1*, showed strong evidence for association with IL-1β ($P = 0.000016$), IL-6 ($P = 0.0005$) and TNF-α ($P = 0.00089$).

To predict statistically the causal variants responsible for the observed associations, we carried out Bayesian quantitative trait nucleotide (BQTN)[16] analysis to examine all 8,192 possible additive gene action models. BQTN analysis provided overwhelming support for a model in which a single SNP (−105G→A) was responsible for the observed associations, with estimated posterior probabilities of >0.999, 0.95 and 0.79 for TNF-α, IL-1β and IL-6, respectively, that the −105G→A promoter polymorphism was of direct functional consequence (or was highly correlated with a functional variant). These high posterior probabilities of effect strongly suggested that the −105G→A variant was required to predict optimally the plasma levels of proinflammatory cytokines in this data set. Similarly, the BQTN analysis provided experiment-wide P values of 0.000038 for the effect of the −105G→A polymorphism on TNF-α levels, 0.0011 for IL-1β levels and 0.0496 for IL-6 levels. Therefore, this promoter SNP significantly influences plasma levels of all three proinflammatory cytokines. Variance component calculations indicated that >6.5% of the variation in IL-1β levels, 5.7% of the variation in TNF-α levels and 3.6% of the variation in IL-6 levels was attributable to variation in *SEPS1*. This is more than twice as much variation in these traits than is accounted for by sex and age.

The effect of the −105G→A polymorphism on mean cytokine levels was large and conformed to additive gene action. GA heterozygotes had mean IL-1β levels that were 0.533 ± 0.133 standard deviation (s.d.) units higher ($P = 0.00006$) than those of common GG homozygotes, whereas rare AA homozygotes had mean levels that were 1.125 ± 0.493 s.d. units higher ($P = 0.0225$). This represents a large effect on individual propensity for inflammation. Similarly, GA heterozygotes had mean TNF-α levels that were 0.474 ± 0.131 s.d. units higher ($P = 0.0003$) than those of GG homozygotes, and AA homozygotes had mean levels that were 1.000 ± 0.489 s.d. units higher ($P = 0.0409$). The effect of the A allele was somewhat smaller on IL-6 levels: GA heterozygotes had mean levels that were 0.431 ± 0.128 s.d. units higher ($P = 0.0007$) than those of GG homozygotes, and AA homozygotes had mean levels that were 0.679 ± 0.441 s.d. units higher ($P = 0.12$). In all cases, formal tests constraining the genotype-specific mean for the heterozygous genotype to a position halfway between the means for the two homozygous genotypes were not significantly different from models in which three unconstrained genotypic means were estimated.

Because of the high level of genetic correlation among these three cytokine measures, we also carried out a multivariate association analysis using SOLAR. This simultaneous analysis of the three inflammation phenotypes showed a highly significant association between the −105G→A polymorphism and plasma cytokine levels ($P = 2.0 \times 10^{-7}$).

We also carried out association analyses on predicted *SEPS1* haplotypes. For each individual, we obtained the most probable haplotypes using the program MERLIN[13]. We also carried out BQTN analysis on the haplotype data, but haplotypic information did not improve the association. In general, the results of haplotype analysis were weaker owing to the increased number of tests incurred and the imperfect correlation between the −105G→A polymorphism and each haplotype.

Because we compiled the study sample by identifying obese probands, we considered whether the association between the −105G→A polymorphism and proinflammatory cytokines could be mediated by

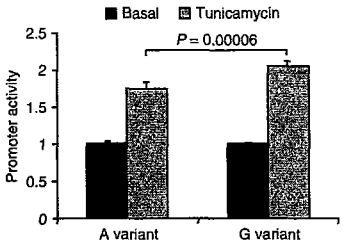

Figure 4 Effect of *SEPS1* SNP −105G→A on expression of *SEPS1* after challenging HepG2 cells with tunicamycin. Comparison of promoter activities reflects the difference between basal luciferase activity and luciferase activity in samples treated with tunicamycin.

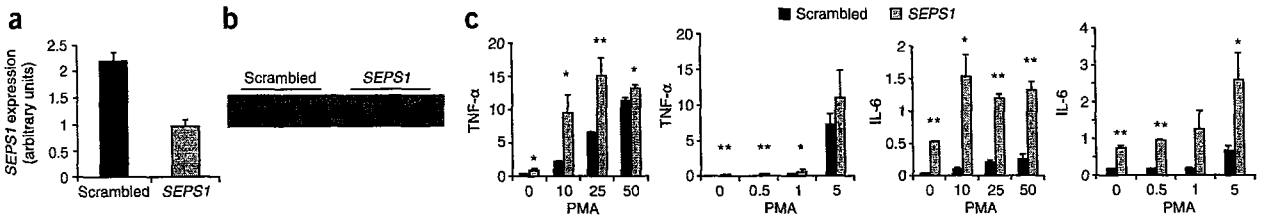

Figure 5 siRNA suppression of SEPS1 in macrophages. (**a**,**b**) Effects of siRNA suppression on SEPS1 mRNA (**a**) and protein (**b**) levels in RAW264.7 macrophage cells. (**c**) Cells were treated with phorbol myristate acetate (PMA) or phytohemagglutinin (PHA) at the indicated concentrations (μg ml^{-1}), and IL-6 and TNF-α levels (ng per mg protein) in the media were measured by ELISA. $^*P < 0.05$ and $^{**}P < 0.005$ for comparison of cells treated with *SEPS1* siRNA versus cells treated with scrambled siRNA.

obesity. Simultaneous correction for the effect of body mass index (BMI) on plasma cytokines did not materially change our results, and we found no evidence for polymorphism-BMI interaction. The −105 G→A polymorphism remained strongly associated with all three inflammation markers ($P = 0.00047$ for IL-1β, $P = 0.000013$ for IL-6 and $P = 0.000086$ for TNF-α), even after we removed the effect of obesity. Therefore, the observed associations are independent of obesity.

Functional analysis of the −105G→A promoter polymorphism

Because the results of the BQTN analysis suggested that the −105G→A promoter polymorphism was either directly functional or in high linkage disequilibrium with a functional variant that was not detected in our resequencing, we decided to test directly the functionality of this variant in relation to its possible effect on *SEPS1* expression. The *SEPS1* proximal promoter is GC-rich and contains one conserved and one putative ER stress-response element (ERSE)[17]. The −105G→A polymorphism is located in the center of the putative ERSE. The ER stress inducer tunicamycin (a protein glycosylation inhibitor) induced *SEPS1* gene expression severalfold in HepG2 cells. Tunicamycin also induced a 301-bp promoter fragment (−262 to +39 bp) of *SEPS1* to a similar magnitude as the endogenous gene. Substitution of the A allele for the G allele at position −105, however, substantially reduced promoter activity in HepG2 cells stressed by tunicamycin, indicating that optimal ER stress inducibility of *SEPS1* requires the G allele (**Fig. 4**). The negative control, a vector with no promoter inserted, had negligible activity (data not shown). Formal statistical analysis of these functional analyses showed that cell lines with the A allele had significantly lower expression *SEPS1* ($P = 0.000061$). Therefore, these *in vitro* experiments provide strong evidence that the −105G→A polymorphism has a direct functional role in *SEPS1* expression.

Confirmation of the association in an independent sample

Given the strong original association between the −105G→A polymorphism and proinflammatory cytokines and the experimental evidence for its functional relevance, we carried out a confirmatory association study in a completely independent sample of 419 Mexican Americans. We measured TNF-α and IL-1β plasma levels for these individuals using the same method used for the original sample. The estimated heritabilities were very similar to the those observed in the families from Wisconsin: 0.435 ± 0.110 ($P = 7 \times 10^{-7}$) for IL-1β and 0.365 ± 0.110 ($P = 1.5 \times 10^{-5}$) for TNF-α. The A variant was somewhat more common in the Mexican American sample, with an observed allele frequency of 0.21. Using the same additive genetic model, association analyses confirmed that the −105G→A polymorphism had a significant effect on levels of both TNF-α ($P = 0.0049$) and IL-1β ($P = 0.0101$). In both cases, the A allele was associated with higher levels of the proinflammatory cytokines.

Role of *SEPS1* in cytokine production

To assess the role of *SEPS1* in cytokine production, we suppressed *SEPS1* gene expression using short interfering RNA (siRNA) in RAW264.7 macrophage cells in two independent experiments, each with three replicates. After 24 h of siRNA treatment, *SEPS1* was suppressed by ~60% at both the mRNA and protein levels (**Fig. 5**). We then stimulated the cells with either phorbol myristate acetate or phytohemagglutinin and observed a dose-dependent increase in the levels of IL-6 and TNF-α in the media (**Fig. 5**). These results showed that impairment of *SEPS1* was directly associated with increased cellular cytokine production and release.

DISCUSSION

Inflammation has a role in and often precedes the onset of many common complex diseases, including cancer, atherosclerosis, diabetes and Alzheimer disease. The role of genetic factors in inflammation is poorly understood. Previous family-based studies of plasma levels of TNF-α suggested that heritability ranged from 0.26 (ref. 18) to 0.80 (ref. 19). This wide range might reflect differences in measurement methods, handling of plasma samples, covariate adjustment and study design. In contrast, in this study, we estimated heritability for TNF-α level to be very similar in families from Wisconsin and in Mexican Americans (0.44 and 0.37, respectively). In a recent study of an isolated Nepalese population that used the same measurement method, we estimated heritability for plasma TNF-α levels to be 0.46, again consistent with these estimates[20]. For plasma levels of both TNF-α and IL-6, the lower heritabilities observed elsewhere[18] (and the reduced genetic correlations[18]) might be a result of differences in handling the plasma samples, which should optimally be frozen at −80 °C as soon as possible to minimize degradation. For IL-6 plasma levels, previous studies estimated heritability to be between 0.17 (ref. 18) and 0.24 (ref. 19); this contrasts with our higher estimate of 0.37 in Mexican Americans. We found no published estimates of heritability for IL-1β plasma levels, although our estimates in Mexican Americans (0.44) and in families from Wisconsin (0.49) were very consistent. Additional complexities such as genotype-age and genotype-sex interactions might also contribute to variation in heritability estimates, but evaluation of this possibility awaits further analyses. Taken together, our data support the idea that genetics factors have a key role in inflammation. Therefore, the search for specific genes underlying the heritability of plasma levels of proinflammatory plasma cytokines is justified.

Our data suggest that variants in the newly discovered gene *SEPS1* might be a crucial genetic link influencing the proinflammatory cytokine profiles observed in these common disorders. Genetic variation in *SEPS1* was strongly associated with circulating levels of IL-1β, TNF-α and IL-6. One promoter polymorphism (−105G→A) was also

associated with impaired expression of *SEPS1*, and siRNA suppression of *SEPS1* resulted in increased cytokine production and release in RAW264.7 macrophage cells. Taken together, these data provide strong evidence that genetic variation in *SEPS1* could affect *SEPS1* gene expression and SEPS1 protein production, which in turn could substantially increase cytokine production and release from a range of cells in response to cellular stresses.

Our genetic results showing the importance of *SEPS1* in the pro-inflammatory response are unequivocal. We observed strong associations between three polymorphisms (−105G→A, 3705G→A and 5227C→T) in *SEPS1* and plasma levels of all three proinflammatory cytokines, which accounted for 4–7% of the total phenotypic variance in plasma cytokine levels. This level of genetic determination for a sequence variant is substantial and is comparable to that observed for other genes related to complex diseases, such as the relationship between the Factor V Leiden variant and the activated protein C ratio in Southern European populations[21], *APOE* variants and cholesterol levels[22], and the *PPARG* Pro12Ala variant and susceptibility to diabetes[23].

BQTN analysis strongly supported the hypothesis that the promoter polymorphism (−105G→A) was responsible for influencing phenotypic variation in levels of these cytokines. The posterior probability of direct effect (or functionality) for the −105G→A variant ranged from 0.79 to >0.999, depending on the phenotype considered. Such a high posterior probability of effect is suggestive of a direct functional effect for this variant. We then confirmed this statistical prediction using classical *in vitro* functional experiments, which showed that variation at this promoter site led to differential response of *SEPS1* gene expression to ER stress ($P = 0.00006$). A replication study done in an independent sample also confirmed a significant association between the promoter polymorphism and plasma levels of TNF-α and IL-1β, providing further evidence that this SNP influences variations in proinflammatory cytokine levels. The results suggest that *SEPS1* has a role in the ER stress-response pathway and are consistent with recent experimental results showing that SEPS1 protects cells from oxidative stress[17]. These results are also consistent with the recent report that SEPS1 has a role in the removal of stressor-induced misfolded proteins from the ER, preventing the accumulation of these proteins and the subsequent stress response that leads to activation of the inflammation cascade[4].

The molecular basis for the effect of the −105G→A promoter variant seems to involve the disruption of an ERSE. The promoter region of *SEPS1* contains one confirmed ERSE. ERSEs are involved in the stress inducibility of ER proteins, which assist protein folding in the ER and promote cell survival under stress conditions[24]. The *SEPS1* promoter also contains a sequence, ATTGG-N_6-CCACG, that is similar to that of ERSE-II (ATTGG-N-CCACG)[25]. ERSE-I elements have a similar tripartite structure, CCATT(N_9)CCACG. Stress inducibility by ERSE-I elements requires integrity of both binding motifs as well as a conserved GGC motif in the 9-bp region. Mutations in this GGC motif result in loss of stress inducibility[26]. The −105G→A polymorphism occurs in this string of six nucleotides in the *SEPS1* ERSE-II–like element, located in the promoter at positions −82 to −67, upstream of the transcription initiation site (−82attggccGggaccacg−67). This is in the same region as the biologically important conserved motifs of other ERSEs.

We also showed that suppression of *SEPS1* using siRNA resulted in enhanced cytokine production and release in response to stress in a macrophage cell line. These data are suggestive of a direct mechanistic link between genetic variation in *SEPS1* and circulating cytokine concentrations. The −105G→A variant in *SEPS1* reduces expression of this gene in response to cellular stress, reducing the amount of SEPS1 protein available for processing and removing misfolded proteins from the ER. The accumulation of misfolded proteins in the ER causes stress and results in increased production and release of cytokines.

Selenium is essential for mammalian immune function and antioxidant activity, and selenoproteins are responsible for these biologic effects of selenium in mammals[27,28]. We previously showed that one mechanism of *SEPS1* involvement in immune function might be as a receptor for the acute phase reactant serum amyloid A[1]. But recent investigation of the role of *SEPS1* in modulating ER stress suggests that it might have a more direct role in mediating inflammation[4]. We speculate that under conditions of ER stress, such as accumulation of unfolded proteins or ER overload, the induction of *SEPS1* is required to maintain lumen homeostasis suitable for protein folding and post-translational modification.

Our findings highlight *SEPS1* as a new candidate mediator of inflammatory response. Genetic variation in *SEPS1* showed a highly significant association with cytokine levels, and further functional analysis provided mechanistic evidence to support the idea that *SEPS1* has a previously unrecognized role in inflammatory response and might be involved in the pathogenesis of many common complex diseases influenced by inflammation. Further research on *SEPS1* will be required to determine whether this unique pathway is a good candidate for pharmaceutical targeting.

METHODS

Subject selection. We used human samples and phenotypic data from the Metabolic Risk Complications of Obesity Genes project. The samples are discussed at length elsewhere[29]. These families were recruited from the membership of Take Off Pounds Sensibly, Inc. The company provided mailing material on membership of chapters in ten US states (Wisconsin, Illinois, Michigan, Iowa, Minnesota, Ohio, West Virginia, Missouri, Kentucky and Indiana). Families with at least two obese siblings (BMI >30 kg m^{-2}), with one or both parents available and with at least one sibling or parent who was never obese (BMI <27 kg m^{-2}) were identified and contacted. Families then visited satellite centers (four to six per state), where informed consent was obtained from all individuals and an experienced team undertook the phenotypic procedures. All protocols were approved by the Institutional Review Board of the Medical College of Wisconsin. The data used in the present examination are from 522 individuals in the largest 92 families of predominantly northern European ancestry and residing in the US.

Sequencing sample selection. For variant identification, we selected 50 individuals from three different ethnic populations. The resequencing cohort was made up of ten individuals from the Metabolic Risk Complications of Obesity Genes project discussed above. The other 40 DNA samples came from the populations of Mauritius and Nauru. These samples included 34 obese, diabetic or obese and diabetic individuals and 6 'hypernormal' individuals from each population (age range = 50–70 y), all of whom had measured physical and biochemical parameters within the normal physiological range. Research procedures for this study were approved by the International Diabetes Institute, and informed consent was obtained from all subjects.

Replication study sample selection. For the replication study, we obtained samples and phenotype information from the San Antonio Family Heart Study[30]. The sample consisted of 419 Mexican American individuals from 23 families. These families were randomly selected without regard to phenotype. There were 163 males and 256 females in this sample with an age range of 23–89 y. All protocols were approved by the Institutional Review Board of the University of Texas Health Science Center at San Antonio, and informed consent contained from all subjects.

Phenotyping. Because our previous studies of *SEPS1* suggested that it had a role in the inflammation pathway[1], we measured plasma levels of three proinflammatory cytokines, IL-1β, IL-6 and TNF-α, to test whether genetic

variation in *SEPS1* influenced inflammatory response. Phenotyping was done on a Luminex 100 compact analyzer using a multiplex of three different color-coded bead sets (microspheres), produced by Linco. Each bead set was coated with a reagent specific to each cytokine measured, allowing the capture and detection of each cytokine simultaneously from a single sample. Assays were done in 96-well plates on plasma samples stored at −80 °C. All plasma samples were analyzed in a single batch. Intra-assay coefficients of variation were 4.8%, 6.1% and 6.3% for IL-6, TNF-α and IL-1β, respectively.

Primer design. We sequenced a 9.3-kb region encompassing the promoter, exons, introns and sequences identified as conserved between human and mouse. All gene sequences were analyzed for repetitive DNA using RepeatMasker to facilitate primer design. We designed primers, using Primer3, to be 20–30 bp long with annealing temperatures within 1 °C of each other and within the range of 58–63 °C.

Variant identification. We carried out PCR using 5 ng of genomic DNA in a 20-μl reaction with 0.5 Units of *Taq* DNA polymerase (QIAGEN) on a PCR Express thermal cycler (Thermo Hybaid). We purified PCR products using the spin procedure for the *QIAquick 96 PCR Purification Kit* (QIAGEN). Cycle sequencing was done on both sense and antisense DNA strands in 15-μl reactions with 0.75 μl of ABI PRISM Big Dye Terminators v 3.0 Ready Reaction Mix (Applied Biosystems) and 5.25 μl of *half*TERM Dye Terminator Sequencing Reagent (Genetix Ltd), using the Applied Biosystems dye terminator cycle sequencing protocol, on a PCR Express thermal cycler. We purified sequencing products using *gen*CLEAN 96-well Dye Terminator Removal Plates (Genetix Ltd), dehydrated them, resuspended them in 10 μl of Hi-Di Formamide and ran them through ABI PRISM 3100 POP-6 polymer in a 50-cm ABI PRISM 3100 capillary array, on an ABI PRISM 3100 Genetic Analyzer (all supplied by Applied Biosystems). We did sequence analysis using ABI PRISM *SeqScape Software Version* 1.1, which allows the analysis of raw data and provides quality values to indicate the confidence of the automated base calls (Applied Biosystems).

Genotyping. We carried out genotyping using the MassARRAY system (Sequenom) as previously described[31]. We designed PCR primers using SpectroDESIGNER to amplify ~100 bp surrounding the variant sites and carried out reactions using 2.5 ng of genomic DNA, 2.5 mM $MgCl_2$, standard concentrations of other PCR reagents and 0.1 Units of HotStar *Taq* DNA polymerase (QIAGEN) in a total reaction volume of 5 μl. For multiplex reactions, we diluted and mixed primers to simplify PCR cocktail assembly. After PCR amplification, we carried out a shrimp alkaline phosphatase incubation to digest unincorporated dNTPs and PCR primers in the mix. Using the PCR product as a template, we carried out MassExtend reactions using the flanking extend primers, dideoxynucleotides and a thermostable polymerase to extend the PCR product through the polymorphic site. Finally, we added an ion-exchange resin to purify the product of any excess ions that could interfere with the mass spectrometry analysis.

For the SNP assays that failed primer extension assay design on the MassARRAY system (−105G→A), custom TaqMan SNP genotyping assays were designed by Applied Biosystems. We built a submission file containing the SNP sequence with the File Builder software and sent it to Applied Biosystems by email. They designed and tested assays and sent them back to us, ready for typing. We carried out PCR in a 10-μl final volume, consisting of 5 μl of 2× TaqMan Universal PCR Master mix, 0.125 μl of SNP Genotyping assay mix and 10 ng of DNA. PCR was done on the ABI PRISM 7700 with cycling conditions of 95 °C for 10 min, followed by 55 cycles of 92 °C for 15 s and 60 °C for 60 s. For reading the plate, we used a film seal and set the instrument to use a 13-s reading time.

To ensure genotyping quality control, we included multiple positive and negative control samples in all genotyping plates. When we obtained unexpected results for the controls, we ignored all data from the plate and repeated the reactions from step 1. We transferred all DNA samples from 96-well stock plates to 384-well working plates using a 96-head Beckmann Multimek robotic system to minimize pipetting errors and to help eliminate sample plating errors.

Cell culture and reporter assays. We cultured hepatoma HepG2 cells in monolayer in Dulbecco's modified Eagle medium (DMEM; with 4.5 g l^{-1} of glucose) supplemented with 2 mM glutamine, 1 mM pyruvate, 10% (v/v) heat-inactivated fetal bovine serum in 5% CO_2 and 95% air at 37 °C. We amplified a 301-bp region of the *SEPS1* promoter (−262 bp to +39 bp) using the GC-2 PCR kit (BD Biosciences) from human genomic DNA containing the −150A or −150G variant as templates. We ligated the PCR products into the pGL3 basic firefly luciferase vector (Promega) and confirmed sequences by DNA sequencing. As a negative control, we used a vector with no promoter inserted. We cotransfected HepG2 cells with the reporter construct and a control plasmid, pRL-SV40, that contained the *Renilla* luciferase gene. Twenty-four hours after transfection, we treated cells with tunicamycin (10 μg ml^{-1}) in serum-free DMEM for 24 h to induce ER stress. We assayed firefly and *Renilla* luciferase activities in cell lysates using the Dual-Luciferase system (Promega). Promoter activity is expressed as the ratio of firefly luciferase to *Renilla* luciferase activities, with the ratio in control cells (without tunicamycin treatment) arbitrarily defined to be 1.

siRNA studies. We cultured macrophage (RAW264.7) cells in DMEM with 10% fetal bovine serum. We synthesized siRNAs for SEPS1 and the scrambled control *in vitro* using a kit from Ambion. Primer sequences are given in **Supplementary Table 3** online. We transfected RAW264.7 cells with siRNA oligonucleotides (20 nM) using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, we collected cells for isolated of total RNA or protein or treated cells with phorbol myristate acetate or phytohemagglutinin at the indicated concentrations in DMEM for 8 h to analyze cytokine production. We quantified *SEPS1* mRNA using real-time RT-PCR technology (sequences given in **Supplementary Table 3**). We analyzed SEPS1 protein by western blotting using the SEPS1-specific polyclonal antibody described previously[2]. We analyzed cytokines (TNFα, IL-1β and IL-6) in the media using ELISA kits (BD Biosciences). We normalized cytokine content against cellular soluble protein content.

Statistical genetic methods: multivariate quantitative genetic analysis. To estimate the heritabilities and genetic correlations among the three inflammation phenotypes, we used the computer program SOLAR to carry out a multivariate variance components analysis[14]. Multivariate genetic analysis in SOLAR permits unbalanced data structures (*i.e.*, an individual's data is included in the analysis even if measures are available only on a partial set of phenotypes) so that all available phenotypic information is used. We used all pedigree information simultaneously. We used likelihood ratio tests to test the hypothesis that each phenotype had a significant genetic component and that any given pair of phenotypes shared a common genetic determination.

Quantitative trait nucleotide (QTN) analysis. Given complete sequence data for a gene with functional sites, we can identify statistically which polymorphism or polymorphisms are most likely to direct affect a phenotype of interest. This activity takes us from the quantitative trait locus to the responsible nucleotide differences (the QTNs[32,33]) that influence the phenotype. Although determination of the mechanism by which a genetic variant leads to phenotypic variation will still require molecular investigation, it is possible to formulate a first-line statistical genetic approach to limit the number of genetic variants that need to be examined in the molecular laboratory and to prioritize them in terms of their probable importance in the population.

The QTN model. The QTN model that we used is a simple extension of the classical variance component model[16]. Assume that the candidate locus has m polymorphic nucleotide sites. Define a variable s_i for the ith SNP that takes the values of −1, 0 and 1 for the marker genotypes AA, Aa and aa, respectively. In general, the additive genetic variance associated with the ith marker (σ^2_{ai}) is given by $H_i \times \alpha_i^2$, where H_i is the heterozygosity and α_i is twice the displacement between the homozygous marker means. If the ith locus is nonfunctional but is associated with the phenotype because of linkage disequilibrium with the jth marker, which is a functional variant, then $\sigma^2_{ai} = \rho^2_{ij} \times H_j \times \alpha_j^2 = \rho^2_{ij} \times \sigma^2_{aj}$, where ρ is the correlation between the variables s_i and s_j. The term ρ_{ij} is also the correlation between the allelic values of the two loci and is therefore one of the standard measures of linkage disequilibrium. Note that $\sigma^2_{ai} \leq \sigma^2_{aj}$ (*i.e.*, the variance associated with a marker is generally less than

that due to the functional polymorphism) unless the genotypes at the two loci are completely correlated. Using this framework, we model the phenotype, p, as a linear combination of fixed effects and random variables: $p = \mu + \Sigma\alpha_i s_i + \Sigma\beta_l x_l + \Sigma q_k + g + e$, where μ is the trait mean; β_l is the fixed-effect regression coefficient for a measured covariate (x_l); and q_k, g and e are random effects representing other quantitative trait loci, residual genetic effects and random environmental effects, respectively. Estimation of the various fixed effects and variance components associated with the random effects can be done using standard maximum likelihood methods.

Model selection using the Bayesian information criterion (BIC). Once the extensive polymorphism in a positional candidate gene is assayed, Bayesian model averaging and model selection can be used to determine which polymorphisms are most likely to be directly functional. If all functional variants have not been captured by resequencing, however, the procedure will choose the polymorphisms most highly correlated with the nontyped functional variants. Because there may be a large number of SNPs to evaluate in a candidate gene, there can be many possible models of QTN action. If we consider only additive QTN effects, there are 2^m possible models, where m is the number of QTNs considered. Our approach is to evaluate all such models and to use Bayesian methods to estimate the probability that each SNP has a direct effect on the phenotype.

In a Bayesian framework, comparison of two competing hypotheses can be done by the evaluation of the Bayes factor, which is the ratio of the integrated likelihoods of the competing models[34]. For QTN analyses using this approach, the BIC is defined with reference to the null model. In the null model, there are no fixed QTN effects, but random genetic effects (such as polygenic effects) are allowed to account for nonindependence in families. The BIC of the kth QTN model is given by $BIC_k = -\Lambda_{k0} + d.f._k \ln N_e$ where Λ_{k0} is the likelihood ratio test statistic comparing the QTN model with the null model, $d.f._k$ is the degrees of freedom for the comparison and N_e is the effective sample size.

The BIC can be used to assess whether the QTN model explains sufficient variation in the phenotype to justify the number of parameters used. BIC differences greater than two are indicative of positive evidence of support for one model over another with posterior probabilities of >75%. Similarly, BIC differences of six units represent strong support favoring a model with 95% posterior probabilities, and BIC differences greater than ten units indicate posterior probabilities of >99% and thus represent very strong support.

Bayesian model averaging in QTN analysis. The BIC can also be used to formulate a simple model averaging approach to estimation that explicitly allows for model uncertainty[35]. Let Y indicate all the data including both phenotypic and genotypic information, and let M_k indicate the kth model. It can be shown that $p(Y|M_k)$ is proportional to $\exp(-(1/2)\ BIC_k)$. Therefore, the posterior probability of the model conditional on the data can be approximated by $p(M_k|Y) \approx [\exp(-(1/2)\ BIC_k)]/[\sum\exp(-(1/2)\ BIC_i)]$. Using this relationship and placing it in the context of QTN analysis, the posterior probability that $\alpha_i \neq 0$ is given by $\sum_{Ki} p(M_k|Y)$, where K_i denotes the set of models for which $\alpha_i \neq 0$. This is the posterior probability that the ith SNP has a direct effect (assuming that all genetic variation has been assayed in the candidate locus). Using the posterior probabilities as weights, the posterior mean and variance of any parameter are easily calculated and can then be used to form a statistical test that will yield the correct experiment-wide P value. Thus, this approach directly accounts for model uncertainty and provides an estimate of our faith that a given SNP had a direct effect on the phenotype. We incorporated the Bayesian model averaging and model selection procedures for QTN analysis into our computer package, SOLAR.

URLs. RepeatMasker is available at http://ftp.genome.washington.edu/RM/RepeatMasker.html. Primer3 is available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html. SOLAR is available at http://www.sfbr.org/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health. Take Off Pounds Sensibly, Inc. provided funds for establishment of the family database and clinical phenotyping. This work was also supported by grants from the National Center for Research Resources to the General Clinical Research Centers at the Medical College of Wisconsin and the University of Texas Health Science Center San Antonio. The statistical genetics computer package, SOLAR, is supported by a grant from the National Institutes of Mental Health. The supercomputing facilities used for this work at the SBC Genetics Computing Center were supported in part by a gift from the SBC Foundation. Funds for resequencing, genotyping and functional and statistical analyses were provided by ChemGenex Pharmaceuticals Ltd., Australia.

AUTHORS' CONTRIBUTIONS

J.B., A.H.K., P.Z. and G.R.C. initiated the study. *SEPS1* was initially discovered by G.R.C., K.R.W. and P.Z. J.E.C. and J.B.M.J. supervised the resequencing and all marker typing, aided by J.W. and K.G. J.B.M.J. carried out statistical analyses for the promoter experiments. K.S.E. carried out all required Bioinformatic analyses. Y.G., K.R.W. and G.R.C. were responsible for the functional experiments including the promoter and siRNA components. A.H.K. and D.M.A.A. were responsible for the Wisconsin family samples. J.B., J.W.M., M.C.M. and A.G.C. were responsible for the Mexican American family samples. J.B. was responsible for the cytokine assays in both sets of samples. J.B. carried out or supervised all statistical genetic analyses and was aided by G.C., M.C.M., A.G.C. and T.D.D.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Walder, K. *et al.* Tanis: a link between type 2 diabetes and inflammation? *Diabetes* **51**, 1859–1866 (2002).
2. Gao, Y. *et al.* Elevation in Tanis expression alters glucose metabolism and insulin sensitivity in H4IIE cells. *Diabetes* **52**, 929–934 (2003).
3. Kryukov, G.V. *et al.* Characterization of mammalian selenoproteomes. *Science* **300**, 1439–1443 (2003).
4. Ye, Y., Shibata, Y., Yun, C., Ron, D. & Rapoport, T.A. A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. *Nature* **429**, 841–847 (2004).
5. Karlsson, H.K., Tsuchida, H., Lake, S., Koistinen, H.A. & Krook, A. Relationship between serum amyloid A level and *Tanis/SelS* mRNA expression in skeletal muscle and adipose tissue from healthy and type 2 diabetic subjects. *Diabetes* **53**, 1424–1428 (2004).
6. Pahl, H.L. & Baeuerle, P.A. The ER-overload response: activation of NF-kappa B. *Trends Biochem. Sci.* **22**, 63–67 (1997).
7. Zamani, M., Pociot, F., Raeymaekers, P., Nerup, J. & Cassiman, J.J. Linkage of type I diabetes to 15q26 (IDDM3) in the Danish population. *Hum. Genet.* **98**, 491–496 (1996).
8. Field, L.L., Tobias, R. & Magnus, T. A locus on chromosome 15q26 (IDDM3) produces susceptibility to insulin-dependent diabetes mellitus. *Nat. Genet.* **8**, 189–194 (1994).
9. Blacker, D. *et al.* Results of a high-resolution genome screen of 437 Alzheimer's disease families. *Hum. Mol. Genet.* **12**, 23–32 (2003).
10. Susi, M., Holopainen, P., Mustalahti, K., Maki, M. & Partanen, J. Candidate gene region 15q26 and genetic susceptibility to coeliac disease in Finnish families. *Scand. J. Gastroenterol.* **36**, 372–374 (2001).
11. Saadeddin, S.M., Habbab, M.A. & Ferns, G.A. Markers of inflammation and coronary artery disease. *Med. Sci. Monit.* **8**, RA5–12 (2002).
12. Cheverud, J.M. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* **87**, 52–58 (2001).
13. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
14. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).
15. Abecasis, G.R., Cookson, W.O. & Cardon, L.R. Pedigree tests of transmission disequilibrium. *Eur. J. Hum. Genet.* **8**, 545–551 (2000).
16. Blangero, J. *et al.* Quantitative trait nucleotide analysis using Bayesian model selection. *Hum. Biol.* (in the press).
17. Gao, Y. *et al.* Regulation of the selenoprotein SelS by glucose deprivation and endoplasmic reticulum stress - SelS is a novel glucose-regulated protein. *FEBS Lett.* **563**, 185–190 (2004).
18. de Maat, M.P. *et al.* Genetic influence on inflammation variables in the elderly. *Arterioscler. Thromb. Vasc. Biol.* **24**, 2168–2173 (2004).
19. Pantsulaia, I., Trofimov, S., Kobyliansky, E. & Livshits, G. Genetic and environmental influences on IL-6 and TNF-alpha plasma levels in apparently healthy general population. *Cytokine* **19**, 138–146 (2002).

20. Williams-Blangero, S. *et al.* Genetic influences on plasma cytokine variation in a parasitized population. *Hum. Biol.* **76**, 515–525 (2004).
21. Blangero, J., Williams, J.T. & Almasy, L. Novel family-based approaches to genetic risk in thrombosis. *J. Thromb. Haemost.* **1**, 1391–1397 (2003).
22. Stengard, J.H. *et al.* Contributions of 18 additional DNA sequence variations in the gene encoding apolipoprotein E to explaining variation in quantitative measures of lipid metabolism. *Am. J. Hum. Genet.* **71**, 501–517 (2002).
23. Altshuler, D. *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
24. Caspersen, C., Pedersen, P.S. & Treiman, M. The sarco/endoplasmic reticulum calcium-ATPase 2b is an endoplasmic reticulum stress-inducible protein. *J. Biol. Chem.* **275**, 22363–22372 (2000).
25. Kokame, K., Kato, H. & Miyata, T. Identification of ERSE-II, a new cis-acting element responsible for the ATF6-dependent mammalian unfolded protein response. *J. Biol. Chem.* **276**, 9199–9205 (2001).
26. Roy, B. & Lee, A.S. The mammalian endoplasmic reticulum stress response element consists of an evolutionarily conserved tripartite structure and interacts with a novel stress-inducible complex. *Nucleic Acids Res.* **27**, 1437–1443 (1999).
27. Arthur, J.R. Selenium supplementation: does soil supplementation help and why? *Proc. Nutr. Soc.* **62**, 393–397 (2003).
28. Ferencik, M. & Ebringer, L. Modulatory effects of selenium and zinc on the immune system. *Folia Microbiol. (Praha)* **48**, 417–426 (2003).
29. Kissebah, A.H. *et al.* Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc. Natl. Acad. Sci. USA* **97**, 14478–14483 (2000).
30. MacCluer, J.W. *et al.* Genetics of atherosclerosis risk factors in Mexican Americans. *Nutr. Rev.* **57**, S59–S65 (1999).
31. Buetow, K.H. *et al.* High-throughput development and characterization of a genome-wide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* **98**, 581–584 (2001).
32. Phillips, P.C. From complex traits to complex alleles. *Trends Genet.* **15**, 6–8 (1999).
33. Long, A.D., Lyman, R.F., Langley, C.H. & Mackay, T.F. Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**, 999–1017 (1998).
34. Kass, R.E. & Raftery, A.E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
35. Raftery, A.E. Bayesian model selection in social research. in *Sociological Methodology* (ed. Marsden, P.V.) 111–195 (Blackwells, Oxford, 1995).