

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): E. M. Smith, Y. Zhang, T. M Baye, S. Gawrieh, R. Cole, J. Blangero, M. A. Carless, J. E. Curran, T. D. Dyer, L. J. Abraham, E. K. Moses, A. H. Kissebah, L. J. Martin and M. Olivier

Article Title: INSIG1 influences obesity-related hypertriglyceridemia in humans

Year of publication: 2010

Link to published article: <http://dx.doi.org/10.1194/jlr.M001404>

Publisher statement: This research was originally published in The Journal of Lipid Research. E. M. Smith et al. INSIG1 influences obesity-related hypertriglyceridemia in humans. Journal of Lipid Research. 2010. Vol. 51, pp. 701-708. © the American Society for Biochemistry and Molecular Biology

***INSIG1* influences obesity-related hypertriglyceridemia in humans**

E.M. Smith^{1,2,8*}, Y. Zhang^{1,2*}, T.M. Baye^{1,2}, S. Gawrieh³, R. Cole^{1,2}, J. Blangero⁴, M.A. Carless⁴, J.E. Curran⁴, T.D. Dyer⁴, L.J. Abraham⁵, E.K. Moses⁴, A.H. Kissebah^{1,3}, L.J. Martin⁶, and M. Olivier^{1,2,7#}

¹ Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI, USA

² Biotechnology and Bioengineering Center, Medical College of Wisconsin, Milwaukee, WI, USA

³ Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

⁴ Southwest Foundation for Biomedical Research, San Antonio, TX, USA

⁵ Biochemistry & Molecular Biology, M310 School of Biomedical, Biomolecular & Chemical Sciences, University of Western Australia, 35 Stirling Highway, Crawley, W.A. 6009, Australia

⁶ Cincinnati Children's Hospital, Cincinnati, OH, USA

⁷ Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA

* These authors contributed equally to this project

⁸ Present address: Dept. of Biological Sciences, University of Warwick, Coventry, CV4 7AL, UK

Running Title: *INSIG1* affects human triglyceride levels

corresponding author:

Michael Olivier, Ph.D
Associate Professor
Biotechnology and Bioengineering Center
TBRC, C2910
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI, 53226
USA

phone: ++1 414-955-4968

fax: ++1 414-955-6568

e-mail: molivier@mcw.edu

Abstract

In our analysis of a QTL for plasma TG levels (LOD=3.6) on human chromosome 7q36, we examined 29 single nucleotide polymorphisms (SNPs) across *INSIG1*, a biological candidate gene in the region. Insulin-induced genes (*INSIGs*) are feedback mediators of cholesterol and fatty acid synthesis in animals, but their role in human lipid regulation is unclear.

In our cohort, the *INSIG1* promoter SNP rs2721 was associated with TG levels ($p=2 \times 10^{-3}$ in 1,560 individuals of the original linkage cohort, $p=8 \times 10^{-4}$ in 920 unrelated individuals of the replication cohort, combined $p=9.9 \times 10^{-6}$). Individuals homozygous for the T allele had 9% higher TG levels, and two-fold lower expression of *INSIG1* in surgical liver biopsy samples when compared to individuals homozygous for the G allele. Also, the T allele showed additional binding of nuclear proteins from HepG2 liver cells in gel shift assays. Finally, the variant rs7566605 in *INSIG2*, the only homologue of *INSIG1*, enhances the effect of rs2721 ($p=0.00117$). The variant rs2721 alone explains 5.4% of the observed linkage in our cohort, suggesting that additional so far undiscovered genes and sequence variants in the QTL interval also contribute to alterations in TG levels in humans.

Key words: triglyceride, *INSIG2*, SNP association, gene expression, EMSA

Introduction

Increased plasma triglyceride (TG) levels are an important cardiovascular risk factor, and are strongly associated with atherosclerotic heart disease (1,2). Plasma TG levels vary widely between individuals, and both genetic and environmental factors have been shown to contribute to elevated plasma TG concentrations (3-7). Elevated plasma TG levels are often observed in obese and diabetic individuals and in individuals affected by the metabolic syndrome, a common chronic disorder associated with obesity, insulin resistance, hypertension, and alterations in plasma lipid profile such as elevated serum triglycerides and low HDL levels. This characteristic pattern is similar to lipid abnormalities reported in familial combined hyperlipidemia (8).

The Metabolic Risk Complications of Obesity Genes (MRC-OB) project was established in 1994 to identify the genetic determinants of the metabolic syndrome and its metabolic abnormalities (9). As part of the project, 2,209 individuals from 507 families were ascertained for basic anthropomorphic phenotypes, plasma lipid measures, and fasting glucose and insulin levels. A genome-wide linkage scan of these families identified a quantitative trait locus (QTL) on human chromosome 7q36 linked to plasma triglyceride levels (LOD=3.7) (10). This region has also been implicated in numerous other studies (11-15), and represents one of the most replicated linkages for dyslipidemia and elevated TG levels. The genomic interval on chromosome 7q36 includes 22 known genes, including insulin-induced gene 1 (*INSIG1*), the only biological candidate.

INSIG1 was initially described as an insulin-induced transcript, designated CL-6 (16). It was later re-named *INSIG1* following cloning and assignment to chromosome

7q36 (17). *INSIG1* has been shown to affect the feedback regulation of cellular cholesterol (18, 19). Sterol regulatory element binding proteins (SREBPs), which are transcription factors located in the endoplasmic reticulum, control the synthesis of cholesterol and fatty acids. When cellular cholesterol levels fall, SREBPs bind to SREBP cleavage-activating proteins (SCAPs) in order to be transported to the Golgi, where they undergo proteolytic cleavage (20). The NH₂-terminal fragment translocates to the nucleus to activate the transcription of target genes including *INSIG1* (21) that are involved in the synthesis of cholesterol and fatty acids. As cellular sterol levels rise, SCAP binds to *INSIG1*, thus preventing the transport of the SCAP/SREBP complex to the Golgi (18) and halting the synthesis of cholesterol and fatty acids (for detailed reviews, see (22, 23)). *INSIG1* and its only mammalian homologue *INSIG2* may also exert their feedback regulators of cholesterol synthesis by directly binding to 3-Hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase, which triggers sterol-accelerated ubiquitination and degradation of HMG CoA reductase, the rate-limiting enzyme of cholesterologenesis (for a detailed review, see (24)).

Over-expression of *INSIG1* has been shown to reduce high levels of TG in both liver and plasma of Zucker diabetic fatty rats (25). It has been proposed that this effect is due to the mechanism outlined above (18). In addition, single knockout of either *INSIG1* or *INSIG2* in mice leads to increased total content of both cholesterol and TG in mouse livers, an effect exacerbated significantly when both *INSIG* genes are knocked out (26). These studies indicate that *INSIG1* and *INSIG2* interactively contribute to the control and regulation of lipid homeostasis in animals.

Based on this evidence, we hypothesized that sequence variation in *INSIG1*, which is located in the QTL region on human chromosome 7q36 linked to plasma TG levels, contributes to the observed genetic effect on plasma lipid levels. Here we present data on our comprehensive association analysis of sequence variants in *INSIG1* in the initial MRC-OB cohort and a second replication cohort of unrelated individuals, and evidence for a synergistic effect of an associated sequence variant in *INSIG1* with a variant in *INSIG2* on plasma TG levels. Furthermore, our data suggest that a putative causal promoter variant in *INSIG1* affects gene expression of *INSIG1* in liver tissue, potentially by altering the binding of nuclear factors to the promoter sequence.

Methods

Study Design

We performed a family-based study on genetic contributions to plasma triglyceride levels.

Study cohort

Cohort recruitment and individual phenotyping have been described in detail previously (9,10). In brief, families with at least two obese siblings [body mass index (BMI) >30], the availability of at least one parent, and one or more never obese siblings (BMI <27) were recruited from the TOPS (Take Off Pounds Sensibly, Inc.) membership in ten Midwestern U.S. states. Health information of all participants was obtained by a questionnaire. Individuals were excluded from recruitment with the following conditions: pregnancy, type 1 diabetes, history of cancer, renal or hepatic disease, severe coronary artery disease, substance abuse, corticosteroids or thyroid dosages above replacement dose, history of weight loss of more than 10% in the preceding 12 months, as well as individuals receiving lipid-lowering medications. Phenotypic measurements included height, weight, waist and hip circumferences, and fasting plasma levels of glucose, insulin, total cholesterol, LDL-cholesterol, HDL-cholesterol, and TG. A total of 2,209 individuals distributed over 507 families of Northern European descent qualified for the above-mentioned criteria and thus formed the initial study population, of which 1,560 individuals from 261 families that contributed to the linkage on chromosome 7q36 (family LOD >0) were selected for further studies reported here.

Confirmatory studies were performed from a cohort of unrelated individuals of Northern European descent recruited by the same criteria from the same geographic regions. Samples from 920 unrelated individuals were available for this study. All protocols were approved by the Institutional Review Board of the Medical College of Wisconsin, and all participants signed an informed consent. Details on the two cohorts used in this study are included in Table 1.

Measurements

Genotyping

SNPs were genotyped using Invader technology as described for previous studies (27-30). Genotyping was carried out in a total volume of 6 μ l, containing 0.5 μ l of PCR product, 0.02 μ l of each primary probe, 0.002 μ l of Invader probe, 1.12 μ l of 2.6M betaine (Sigma, St.Louis, MO), 2.75 μ l TE buffer, 0.35 μ l Cleavase (TWT) and 1.24 μ l FRET mix (TWT). Reactions were denatured at 95°C for five minutes and then heated at 63°C for 15-120 min. Additional SNPs were genotyped on an Affymetrix MegAllele custom-designed 3K array using molecular inversion probe technology, as described previously (28). The location of all SNPs included in the analysis is shown in Figure 1.

Sequencing and SNP identification

A 4,139 bp region was re-sequenced in 47 unrelated individuals selected from the MRC-OB cohort. The region was amplified by PCR using 21 overlapping amplicons 261 to 444 bp in size. PCR products were purified using MultiScreen-FB plates (Millipore); and eluted in sterile H₂O for sequencing. Sequencing was carried out using BigDye v.3.1

chemistry (Applied Biosystems, CA, USA) with standard techniques and analyzed on an Applied Biosystems 3730xl DNA Analyzer. SNP loci were identified by aligning all sequences using POLYPHRED (31), all putative sites were checked manually for confirmation. Results were independently confirmed by aligning all traces with the chromosome 7 reference sequence using the anchored alignment algorithm in POLYBAYES(32) and using POLYPHRED to screen for SNPs.

Expression profiling

RNA isolation, anti-sense RNA generation and Illumina BeadChip hybridization of the liver samples as well as transcript detection and normalization were performed as previously described (33). Briefly, a liver biopsy was obtained from each of 73 morbidly obese subjects of Northern European descent undergoing bariatric surgery. RNA was extracted and analyzed using Illumina Human-6 Expression BeadChips which allow interrogation of almost 48,000 transcripts. Gene expression data for all liver samples were analyzed according to Goring et al. (33). To identify transcripts with detectable quantitative expression in the samples, distribution of expression values for a given transcript were compared to those of the controls imbedded in each chip. All significantly expressed transcripts were identified using an FDR of 0.05. To minimize the influence of overall signal levels, which may reflect RNA quantity and quality rather than a true biological difference between individuals, abundance values of all transcripts were standardized by z-scoring within individuals (using decile percentage bins of transcripts, grouped by average log transformed raw signals across individuals), followed by linear

regression against the individual-specific average log-transformed raw signal and its squared value, as discussed by Goring et al. (33).

For the analysis of allele-specific expression differences in *INSIG1*, individuals were grouped according to genotype (for rs2721), and their gene-specific normalized expression z-scores were compared for genotype-specific differences using ANOVA. Differences between individual genotypes were compared using t-tests.

Electrophoretic Mobility Shift Assay (EMSA)

The human HepG2 cell line was maintained in DMEM, supplemented with 2 mM L-glutamine, 100 mg/ml each of streptomycin and penicillin, and 15% fetal bovine serum, at 37°C with 5% CO₂. Complementary oligonucleotides representing both allelic forms of the *INSIG1* promoter SNP (rs2721, G>T at -2253 to the start codon) were obtained commercially (Integrated DNA Technologies) and purified by high performance liquid chromatography (HPLC). EMSA was performed as described previously (34). Briefly, nuclear extracts were prepared from approximately 8×10⁷ cells. Extracts were frozen in liquid N₂ and stored at -80°C. For EMSA, nuclear proteins (3μg) were pre-incubated for 10 min on ice with 1μg of poly (dI-dC) (Pharmacia) in a binding buffer (4% Ficoll, 20mM HEPES [pH 7.9], 1mM EDTA, 1 mM DTT, 50 mM KCl). Nuclear proteins were then incubated with biotin-labeled double-stranded oligonucleotide probes (20 fmol) for 30 min on ice and then analyzed on a 6% Novex® DNA retardation gels (Invitrogen, CA, USA), and electroblotted onto a positively charged nylon membrane (Ambion, Austin, TX). Detection of protein/DNA complexes was achieved following incubation of the membrane with streptavidin-horseradish peroxidase and development with luminal

substrate (Lightshift Chemiluminescent EMSA kit; Pierce, IL, USA). Light emission was captured on X-ray film.

Statistical Analysis

Initial association analysis discovered rs2721

As triglyceride levels are continuously distributed, the data were first examined for deviations from normality. Raw triglyceride levels exhibited an increased number of high values, so the data were natural log (\ln) transformed. Data were re-examined and observations exceeding 4 standard deviation units were removed as outliers. In all models, triglycerides were adjusted for age, sex, age by sex, age squared, age squared by sex, and type 2 diabetes status.

To test for statistical association with SNPs in *INSIG1*, we identified 11 tagSNPs in the *INSIG1* gene region (~40 kb) from the Hapmap database using the tagger function in Haploview with default settings (35,36). After sequencing the promoter and the first exon of *INSIG1* in 47 unrelated individuals, we identified an additional 17 SNPs. All 28 SNPs were genotyped in the full cohort ($n = 1560$) and were tested for association with $\ln(\text{triglycerides})$.

For each SNP, we used the measured genotype approach to test for association. Briefly, in this approach the effects of the individual SNP genotypes will be modeled by assigning genotypic values such that the homozygotes will be assigned values of 1 and -1 and the heterozygotes will be assigned 0, providing an additive model (37). To account for the phenotypic correlation between family members, we used variance components analysis with the SNPs screened individually as covariates in the computer program SOLAR (Sequential Oligogenic Linkage Analysis Routines) (38).

Confirmatory association analysis in unrelated individuals

To confirm associations identified in the family-based analyses, we genotyped SNP rs2721 in a cohort of 920 unrelated individuals. Genotypes were coded as described above and were considered the independent variable. Triglycerides were ln transformed to improve normality of the distribution. As this was an unrelated cohort, we performed linear regression analysis using SAS v9.1 (SAS, Carey NC).

Meta-Analysis of both cohorts

To assess the combined effect of rs2721 in both cohorts, we combined the p-values using the Z-transformation. This approach is detailed by Whitlock et al. (39) and implemented in the Util program (<http://www.genemapping.cn/util.htm>).

Interaction analysis with INSIG2

To determine whether there was an interaction between *INSIG1* and *INSIG2*, as suggested from mouse knockout studies, we first examined the relationship between the family specific LOD scores for triglycerides at the two loci (40). Second, we genotyped two SNPs in *INSIG2* (rs57566605 and rs2422166) that had been reported to be associated with BMI and LDL-C in other studies (41-43), and tested for both main effects and interaction effects using the measured genotype analysis in SOLAR. Third, we examined the impact of *INSIG1* variants conditional on *INSIG2* variants. These analyses were performed by subsetting the data by *INSIG2* genotype and then performing measured genotype analysis for the effect of *INSIG1* genotype.

Results

Association of *INSIG1* sequence variants with TG in the MRC-OB cohort

The QTL on chromosome 7q35-q36 linked to plasma TG levels has been replicated in multiple cohorts (10-15). Among the 22 identified genes in this QTL, *INSIG1* is the only candidate with known biological roles that are related to lipid metabolism and therefore has been the focus of this study.

Initially, we tested for *INSIG1* association using 11 SNPs selected from the Hapmap database using the tagger function of Haploview (35, 36) (Figure 1 and Table 2). These SNPs spanned the entire *INSIG1* gene region (39.7 kb). One variant in the putative promoter region of *INSIG1*, rs2721, was significantly associated with plasma triglyceride levels in 1,560 individuals from our cohort ($p=0.002$). Two additional SNPs (rs9770068 and rs9769506) located in intron 6 of *INSIG1*, showed marginal association ($p=0.047$ and $p=0.037$, respectively), as well as one additional SNP in the promoter region (rs9690040, $p=0.045$). As these three marginally associated SNPs are in almost complete linkage disequilibrium with rs2721 (Fig.1), we speculate that their high degree of correlation may have led to the observed associations.

The promoter variant (rs2721) is located 2253 bp upstream of the transcription start site of *INSIG1*. In order to include all other variants surrounding rs2721 in the association analysis, we re-sequenced the promoter region and exon 1 in 47 unrelated individuals from our cohort. SNPs in the remaining exons of the gene were in high LD with each other, and not in LD with rs2721, thus, we focused solely on the region surrounding rs2721. All selected individuals were parents from families contributing

significantly to the initially observed linkage on chromosome 7 in the MRC-OB cohort. The re-sequencing effort uncovered 17 additional SNPs, of which 8 were novel and not reported in dbSNP. The location of all SNPs is shown in Figure 1.

All 28 SNPs were genotyped in the 1,560 individuals from the MRC-OB cohort. The results of the association analysis are summarized in Table 2. No additional SNPs were associated with plasma triglyceride levels in this cohort, suggesting that the effect is primarily mediated by rs2721. The TG levels of individuals homozygous for the G allele (123.1 ± 2.3 mg/dl, $n=1073$) were on average 9% lower than those of individuals homozygous for the T allele (135.9 ± 16.2 mg/dl, $n=31$).

Next, we examined the effect of the SNP on the initial linkage. Overall, the QTL interval on chromosome 7q36 explains 31% of the variance in TG levels in our cohort. When including SNP rs2721 in *INSIG1* as a covariate, the linkage drops 5.4% for our family-based cohort, suggesting a modest but significant effect on the initial linkage on chromosome 7q36 that explains 1.6% of the total observed variance in TG levels in the study cohort.

Replication of association analysis

To validate the association seen in the family-based MRC-OB cohort, we examined the effect of the associated variant rs2721 in a second study cohort. The details of the cohort are summarized in Table 1. 920 individuals were genotyped and included in the analysis. The minor allele frequency (MAF) of rs2721 in this cohort is comparable to the MAF of rs2721 in the initial family-based cohort (15.1% vs. 14.5%). Again, rs2721 is strongly

associated with plasma triglyceride levels ($p=8\times 10^{-4}$), confirming the finding in our initial family-based cohort. In this replication cohort, individuals homozygous for the T allele of rs2721 ($n=21$) had mean TG levels of 150.3 ± 18.8 mg/dl, compared to 135.9 ± 2.9 mg/dl for individuals homozygous for the G allele ($n=674$).

When both datasets are combined using a Z-transformation (39), rs2721 is associated with altered TG levels with a p-value of 9.9×10^{-6} .

Functional analysis of rs2721

To assess whether rs2721 may have a direct functional effect, we examined whether the sequence variant is correlated with *INSIG1* expression levels since the SNP is located in the promoter region of the gene. We examined mRNA expression levels of *INSIG1* in 73 liver biopsy samples from morbidly obese patients ($BMI>36$). All individuals were genotyped for rs2721, and average gene expression levels were compared within different genotype groups. The T allele reduced *INSIG1* gene expression levels significantly. In liver, individuals with TT genotype ($n=2$) had 2-fold lower levels of *INSIG1* mRNA when compared to those with TG ($n=23$, $p=0.046$) or GG types ($n=48$, $p=0.030$) (mean z-score difference of 0.52, $p=0.02$). *INSIG1* expression in heterozygous individuals was not significantly different from GG homozygous individuals ($p=0.40$).

Based on this observed correlation between *INSIG1* gene expression and rs2721 genotype, we investigated whether the two alleles may have differential affinities to nuclear proteins such as transcription factors, which may change the transcription levels that eventually lead to the altered lipid profiles. We performed electrophoretic mobility

shift assays (EMSA) using biotin-labeled double-stranded DNA (41 bp) surrounding SNP rs2721 (20 nucleotides both upstream and downstream). As shown in Figure 2, the presence of the T allele in the promoter DNA leads to a shift in observed bands after binding of nuclear extracts from HepG2 cells when compared to the G allele. This shift is not seen in control experiments without nuclear extract or biotinylated probes. These results show that the two alleles of rs2721 have different affinities to nuclear proteins. While the protein(s) binding to the sequence surrounding rs2721 are currently unknown, it is possible that this differential binding mediates the observed difference in *INSIG1* gene expression, and thereby affects plasma TG levels in individuals homozygous for the T allele.

Interaction between *INSIG1* and *INSIG2*

In humans, the *INSIG1* protein is 78% similar to its only homologue *INSIG2*. Interestingly, while *INSIG1* is involved in the feedback regulation of lipid synthesis, *INSIG2* has been shown to be associated with obesity and adipocyte metabolism in a number of recent studies(41-43). However, no study to date has addressed whether these two genes work cooperatively in humans to affect lipid levels, as has been shown in mice (26).

Correlation analysis based on family-specific LOD scores of the locus of *INSIG1* and the locus of *INSIG2* suggested a non-additive interaction between the two genes ($\rho=-0.239$, $p<0.0001$). Based on these results, we selected two SNPs in *INSIG2* (rs2422166 and rs7566605) that have been repeatedly shown to be associated with BMI

and LDL-C in other studies (41-43). Gene-gene interaction analysis was performed on the two *INSIG2* SNPs and rs2721 of *INSIG1*. In this analysis, SNP rs2422166 did not show any effect on TG levels directly or through interaction with rs2721. In contrast, SNP rs7566605 showed marginal interaction with rs2721 ($p=0.0473$) although it was not directly associated with TG levels. To measure the impact of genotype at rs7566605 on the association of rs2721 with lipid traits, we grouped the data based on genotype at rs7566605. Measured genotype analysis was then performed in these subsets (Table 3). In the GG subgroup, the associations with TG levels disappeared. In contrast, in the CC-CG subgroup, the association was significantly strengthened ($\beta=0.13\pm 0.04$, $p=0.00117$). Here, individuals homozygous for the T allele of rs2721 had 31% higher TG levels than individuals homozygous for the G allele (138.4 ± 1.9 mg/dl vs. 105.6 ± 1.7 mg/dl). Before we stratified the study cohort based on genotypes of *INSIG2* rs7566605, the difference between the *INSIG1* rs2721 homozygous groups was only 8%.

Discussion

Studies conducted in animals have revealed intriguing and critical roles of *INSIG* genes in lipid homeostasis (26). By anchoring a sterol-sensing retainer protein that is able to block the access to the nucleus of a number of transcription factors, INSIG proteins have been recognized as key regulators of both cholesterol and fatty acid synthesis (18, 19, 21). No study, however, has shown evidence of their proposed roles in lipid homeostasis in humans. Based on the fact that the gene for *INSIG1* is located in a QTL region linked to plasma TG levels in our MRC-OB cohort, we examined the role of sequence variants in the gene in the modulation of plasma TG levels.

Our initial association analysis using eleven tagSNPs revealed four SNPs associated with TG levels. Further analysis of additional sequence variants uncovered by re-sequencing suggested the promoter SNP rs2721 as the likely causal variant for the observed hyperlipidemia. Genotype effect assessment showed that people homozygous for the T allele at rs2721 had significantly elevated plasma TG levels. The association was replicated in a second independent cohort, detecting a similar effect in individuals homozygous for the T allele. These data suggest that rs2721 in *INSIG1* may be causally associated with TG levels in our linkage region. However, our analysis also suggests that this variant accounts for only a portion of the overall linkage (~5%), suggesting that other variants and genes in the QTL interval also affect plasma TG levels. Overall, variance at rs2721 explains 1.6% of the observed variance in TG levels in our family cohort. This

contribution of this polymorphism, however, may be underestimated since the interaction with other genes and beyond *INSIG2* locus have not been considered in this study.

The SNP rs2721 resides 2,253 bp upstream of the start codon of *INSIG1* in the putative promoter region. To test the genotype-specific effect on the transcription of *INSIG1* in humans, expression profiling was conducted in liver tissue, the putative target tissue for *INSIG1* action on lipid homeostasis in mice. Our analysis revealed that individuals homozygous for the T allele had significantly lower *INSIG1* expression levels, suggesting that reduced expression of *INSIG1* increases circulating plasma TG levels. While the expression analysis revealed only two individuals homozygous for the T allele in our sample set of 73 individuals, the data nonetheless provide suggestive evidence for an effect of this promoter variant on gene repression *in vivo*. Additional samples and tissues will need to be investigated to further substantiate this effect of rs2721.

To further validate this potential effect of SNP rs2721 on the promoter function of *INSIG1*, we examined the binding of nuclear proteins to the sequence surrounding rs2721 by electrophoretic mobility shift assay. Separate custom oligonucleotides specific for the alleles of rs2721 and comprising the immediate sequence surrounding the SNP location were designed. Oligonucleotides specific for the T allele were able to bind nuclear proteins extracted from HepG2 cells, in contrast to the G allele (Figure 2). These data demonstrate that the T allele of rs2721 shows differential binding of nuclear proteins *in vitro*, further supporting the hypothesis that this sequence variant affects promoter

function of *INSIG1*. Analysis of the sequence using TFBIND (<http://tfbind.ims.u-tokyo.ac.jp/>) reveals that rs2721 is located in a putative binding site for the transcription factor Pre-B-cell leukemia homeobox 1 (PBX1) which is expressed in the liver and a wide variety of other tissues. The binding site is altered in the G allele, reducing the probability of PBX1 binding more than 6-fold over the T allele. Additional transcription factors with differential binding affinity to the two alleles of rs2721 include Forkhead transcription factor (FOXO1) and Caudal type homeobox A (CDXA). This differential binding affinity may explain the observed EMSA results, and may be related to the observed allele-specific expression difference, although none of the transcription factors have been reported to modulate plasma triglyceride levels. Nonetheless, PBX1 has been reported to affect glucose tolerance and steroidogenesis, and it has been suggested that mutations in PBX1 may promote susceptibility to diabetes (44). However, at this point, the specific proteins mediating this effect (binding to the T allele, reducing expression of *INSIG1*, and thereby elevating plasma TG levels) remain to be identified. While the EMSA results, in conjunction with the expression analysis of the liver biopsy samples, suggest a functional mechanism for rs2721, it is clear that EMSA results do not prove functional interactions, and additional molecular studies using transfection assays may be required to confirm the initial findings.

It has been shown in animal models that double knockout animals for *INSIG1* and its only homologue *INSIG2* have more significant lipid abnormalities when compared to knockout animals for either gene individually (26). We therefore explored the putative interactive effect of sequence variants in both human *INSIG1* and *INSIG2* on plasma

lipids. By focusing on the functional SNP rs2721 of *INSIG1*, we tested the gene-gene interaction with two *INSIG2* variants that have shown repeatedly to be associated with BMI and lipid abnormalities in obesity studies. Based on our analysis, *INSIG1* does interact with *INSIG2* in relation to lipid homeostasis. The two *INSIG2* SNPs have no independent effect on plasma TG levels in our cohort. However, rs7566605 significantly enhances the effect seen with rs2721 of *INSIG1*, and their interaction strengthened the association with plasma TG levels (Table 3). Specifically, people possessing GG genotype at rs7566605 of *INSIG2* no longer show a detectable effect of the TT genotype of rs2721 in *INSIG1* on TG levels. In contrast, individuals with at least one C allele at rs7566605 of *INSIG2* showed an enhanced effect of the T allele of rs2721 on TG. These data suggest that *INSIG2* (or at least the G allele of rs7566605) may be able to compensate for the effect of rs2721 in *INSIG1*. While the exact mechanism remains to be elucidated, our analysis overall clearly suggests that rs2721 affects plasma TG levels by lowering *INSIG1* gene expression in the liver. This effect can be reduced or enhanced by the different alleles of rs7566605 in *INSIG2*.

Clinically, a reduction of 10 mg/dl in plasma TG levels, independently of the impact of treatment on LDL-cholesterol levels, resulted in a 1.4% reduction in risk of developing coronary heart disease (CHD), in a report by Miller M et al. (45). In our study, the elevation in TG levels in individuals with the TT genotype of rs2721 was 12.8mg/dl and 14.4 mg/dl for families or unrelated individuals, respectively when the population was not stratified by genotypes of *INSIG2* at rs7566605. This effect is further exacerbated by the co-occurrence of the C allele of rs7566605 in *INSIG2* and the TT allele of rs2721, resulting in a 32.8 mg/dl elevation in TG levels.

Despite the replication of the linkage finding on chromosome 7q36, all genome-wide association studies (GWAS) to date exploring lipid traits have failed to detect any significant association in this genomic interval. Neither *INSIG1* nor any other gene in the QTL region has been implicated in any of the lipid trait-related GWAS efforts. Given the low linkage disequilibrium between rs2721 and any of the adjacent SNPs on common GWAS platforms, it is not surprising why the association of rs2721 with plasma TG levels detected in our analysis was not uncovered by GWAS. However, the complete lack of association evidence for the entire genomic interval, despite repeated linkage evidence, needs to be investigated further. This effect has been reported before, and is not specific to this QTL region or lipid traits (46).

The lack of comparable findings in other studies may also be due to differences in the recruitment of the study cohorts. In our study design, we excluded individuals with hepatic disease, severe coronary artery disease, dramatic weight loss in the preceding 12 months, or individuals taking lipid-lowering medications. Given the suggested function of the *INSIG* genes, these criteria may themselves be associated with *INSIG* function, thus may have potential influence on the results. Likewise, it remains to be seen whether the expression profiles obtained from liver samples collected from bariatric surgery patients reflect the levels of *INSIG1* expression in the general population.

While *INSIG1* is not the only gene and rs2721 not the only sequence variant responsible for the initially observed linkage on human chromosome 7q36, our analysis, for the first time, demonstrates a significant role for this gene in the control and

regulation of plasma TG levels, and suggests a functional mechanism by which the causal variant rs2721 may exert its effect. Further QTL-wide analyses will be required to elucidate the other remaining genetic factors affecting plasma TG levels in our cohort, and validate the effect on this important cardiovascular disease risk factor in other patient cohorts.

Figure Legends

Figure 1. Schematic illustration of the *INSIG1* gene region. All SNPs investigated across the gene region (154,696,933 bp-154,736,613 bp) are shown relative to the *INSIG1* gene, depicted by the black line. Translated exons are shown by white boxes, untranslated exonic regions are indicated by grey boxes. SNPs used in the initial association analysis are highlighted in green, SNPs uncovered by resequencing are shown in black. The associated promoter SNP rs2721 is highlighted by the red box. The colored diagram depicts the pairwise linkage disequilibrium between SNPs (D'), as calculated for the family-based cohort of 1,560 individuals using the program Haploview.

Figure 2. Electrophoretic mobility shift assay of rs2721 using nuclear extract of liver HepG2 cells. Oligos of DNA sequences from *INSIG1* promoter region centered with either allele of rs2721 (T or G) were incubated with total nuclear proteins from cultured liver cells. Arrows point at mobility shift of oligonucleotides with the T allele caused by differential protein binding. Results are representative of at least two independent experiments.

Table Legends

Table 1. Summary statistics for the cohorts used in this study, the original 1,560 individuals of the MRC-OB Family Cohort and 920 unrelated individuals of the replication cohort.

Table 2. Summary of SNP association results. All SNPs and their position are listed, together with p-values for association with plasma triglyceride levels in the MRC-OB family cohort and the replication cohort. SNPs highlighted in green were used in the initial association analysis. The results for SNP rs2721 are highlighted by the box.

Table 3. Interaction analysis of rs7566605 in *INSIG2* and rs2721 in *INSIG1*. The C-allele of rs7566605 significantly enhances the association of rs2721 with plasma TG levels.

* = significant difference when compared to GG of rs2721 (p=0.00117)

Acknowledgments

This work was supported by grant HL74168 from the Heart, Lung, and Blood Institute of the National Institutes of Health (M.O.).

References

1. Forrester, J. S. 2001. Triglycerides: risk factor or fellow traveler? *Curr Opin Cardiol.* **16**: 261-264.
2. Malloy, M. J. and J. P. Kane. 2001. A risk factor for atherosclerosis: triglyceride-rich lipoproteins. *Adv Intern Med.* **47**: 111-136.
3. Connelly, P. W., A. Petrasovits, S. Stachenko, D. R. MacLean, J. A. Little, and A. Chockalingam. 1999. Prevalence of high plasma triglyceride combined with low HDL-C levels and its association with smoking, hypertension, obesity, diabetes, sedentariness and LDL-C levels in the Canadian population. Canadian Heart Health Surveys Research Group. *Can J Cardiol.* **15**: 428-433.
4. Tai, E. S., S. C. Emmanuel, S. K. Chew, B. Y. Tan, and C. E. Tan. 1999. Isolated low HDL cholesterol: an insulin-resistant state only in the presence of fasting hypertriglyceridemia. *Diabetes.* **48**: 1088-1092.
5. Austin, M. A., M. C. King, R. D. Bawol, S. B. Hulley, and G. D. Friedman. 1987. Risk factors for coronary heart disease in adult female twins. Genetic heritability and shared environmental influences. *Am J Epidemiol.* **125**: 308-318.
6. Heller, D.A., U. de Faire, N. L. Pedersen, G. Dahlen, and G. E. McClearn, 1993. Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med.* **328**: 1150-1156.
7. Perusse, L., T. Rice, J. P. Despres, J. Bergeron, M. A. Province, J. Gagnon, A. S. Leon, D. C. Rao, J. S. Skinner, J. H. Wilmore, *et al.* 1997. Familial resemblance of plasma lipids, lipoproteins and postheparin lipoprotein and hepatic lipases in the HERITAGE Family Study. *Arterioscler Thromb Vasc Biol.* **17**: 3263-3269.
8. Kissebah, A. H., S. Alfarsi, and P. W. Adams. 1981. Integrated regulation of very low density lipoprotein triglyceride and apolipoprotein-B kinetics in man: normolipemic subjects, familial hypertriglyceridemia and familial combined hyperlipidemia. *Metabolism.* **30**: 856-868.
9. Kissebah, A. H., G. E. Sonnenberg, J. Myklebust, M. Goldstein, K. Broman, R. G. James, J. A. Marks, G. R. Krakower, H. J. Jacob, J. Weber, *et al.* 2000. Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci U S A.* **97**: 14478-14483.
10. Sonnenberg, G. E., G. R. Krakower, L. J. Martin, M. Olivier, A. E. Kwitek, A. G. Comuzzie, J. Blangero, and A. H. Kissebah. 2004 Genetic determinants of obesity-related lipid traits. *J Lipid Res.* **45**: 610-615.
11. Li, W. D., C. Dong, D. Li, C. Garrigan, and R. A. Price. 2005. A genome scan for serum triglyceride in obese nuclear families. *J Lipid Res.* **46**: 432-438.
12. Duggirala, R., J. Blangero, L. Almasy, T. D. Dyer, K. L. Williams, R. J. Leach, P. O'Connell, and M. P. Stern. 2000. A major susceptibility locus influencing plasma triglyceride concentrations is located on chromosome 15q in Mexican Americans. *Am J Hum Genet.* **66**: 1237-1245.
13. Horne, B. D., A. Malhotra, and N. J. Camp. 2003. Comparison of linkage analysis methods for genome-wide scanning of extended pedigrees, with application to the TG/HDL-C ratio in the Framingham Heart Study. *BMC Genet.* **4 Suppl 1**, S93.

14. Lin, J. P. 2003. Genome-wide scan on plasma triglyceride and high density lipoprotein cholesterol levels, accounting for the effects of correlated quantitative phenotypes. *BMC Genet.* **4 Suppl 1**, S47.
15. Shearman, A. M., J. M. Ordovas, L. A. Cupples, E. J. Schaefer, M. D. Harmon, Y. Shao, J. D. Keen, A. L. DeStefano, O. Joost, P. W. Wilson, *et al.* 2000. Evidence for a gene influencing the TG/HDL-C ratio on chromosome 7q32.3-qter: a genome-wide scan in the Framingham study. *Hum Mol Genet.* **9**: 1315-1320.
16. Diamond, R. H., K. Du, V. M. Lee, K. L. Mohn, B. A. Haber, D. S. Tewari, and R. Taub. 1993. Novel delayed-early and highly insulin-induced growth response genes. Identification of HRS, a potential regulator of alternative pre-mRNA splicing. *J Biol Chem.* **268**: 15185-15192.
17. Peng, Y., E. J. Schwarz, M. A. Lazar, A. Genin, N. B. Spinner, and R. Taub. 1997. Cloning, human chromosomal assignment, and adipose and hepatic expression of the CL-6/INSIG1 gene. *Genomics.* **43**: 278-284.
18. Yang, T., P. J. Espenshade, M. E. Wright, D. Yabe, Y. Gong, R. Aebersold, J. L. Goldstein, and M. S. Brown. 2002. Crucial step in cholesterol homeostasis: sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER. *Cell.* **110**: 489-500.
19. Radhakrishnan, A., Y. Ikeda, H. J. Kwon, M. S. Brown, and J. L. Goldstein. 2007. Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: Oxysterols block transport by binding to Insig. *Proc Natl Acad Sci U S A.* **104**: 6511-6518.
20. Brown, M. S. and J. L. Goldstein. 1999. A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proc Natl Acad Sci U S A.* **96**: 11041-11048.
21. Horton, J. D., I. Shimomura, S. Ikemoto, Y. Bashmakov, and R. E. Hammer. 2003. Overexpression of sterol regulatory element-binding protein-1a in mouse adipose tissue produces adipocyte hypertrophy, increased fatty acid secretion, and fatty liver. *J Biol Chem.* **278**: 36652-36660.
22. Horton, J. D., J. L. Goldstein, and M. S. Brown. 2002. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J Clin Invest.* **109**: 1125-1131.
23. Goldstein, J. L., R. A. DeBose-Boyd, and M. S. Brown. 2006. Protein sensors for membrane sterols. *Cell.* **124**: 35-46.
24. DeBose-Boyd, R. A. 2008. Feedback regulation of cholesterol synthesis: sterol-accelerated ubiquitination and degradation of HMG CoA reductase. *Cell Res.* **18**: 609-621
25. Takaishi, K., L. Duplomb, M. Y. Wang, J. Li, and R. H. Unger. 2004. Hepatic insig-1 or -2 overexpression reduces lipogenesis in obese Zucker diabetic fatty rats and in fasted/refed normal rats. *Proc Natl Acad Sci U S A.* **101**: 7106-7111.
26. Engelking, L. J., G. Liang, R. E. Hammer, K. Takaishi, H. Kuriyama, B. M. Evers, W. P. Li, J. D. Horton, J. L. Goldstein, and M. S. Brown. 2005. Schoenheimer effect explained--feedback regulation of cholesterol synthesis in mice mediated by Insig proteins. *J Clin Invest.* **115**: 2489-2498.
27. Olivier, M. 2005. The Invader assay for SNP genotyping. *Mutat Res.* **573**: 103-110.

28. Olivier, M., L. M. Chuang, M. S. Chang, Y. T. Chen, D. Pei, K. Ranade, A. de Witte, J. Allen, N. Tran, D. Curb, *et al.* 2002. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Res.* **30**: e53.
29. Pennacchio, L. A., M. Olivier, J. A. Hubacek, J. C. Cohen, D. R. Cox, J. C. Fruchart, R. M. Krauss, and E. M. Rubin. 2001. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science.* **294**: 169-173.
30. Pennacchio, L. A., M. Olivier, J. A. Hubacek, R. M. Krauss, E. M. Rubin, and J. C. Cohen. 2002. Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. *Hum Mol Genet.* **11**: 3031-3038.
31. Nickerson, D. A., V. O. Tobe, and S. L. Taylor. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* **25**: 2745-2751.
32. Marth, G. T., I. Korf, M. D. Yandell, R. T. Yeh, Z. Gu, H. Zakeri, N. O. Stitzel, L. Hillier, P. Y. Kwok, and W. R. Gish. 1999. A general approach to single-nucleotide polymorphism discovery. *Nature Genetics.* **23**: 452-456.
33. Goring, H. H., J. E. Curran, M. P. Johnson, T. D. Dyer, J. Charlesworth, S. A. Cole, J. B. Jowett, L. J. Abraham, D. L. Rainwater, A. G. Comuzzie, *et al.* 2007. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet.* **39**: 1208-1216.
34. Li, Y. C., J. Ross, J. A. Scheppler, and B. R. Jr. Franza. 1991. An in vitro transcription analysis of early responses of the human immunodeficiency virus type 1 long terminal repeat to different transcriptional activators. *Mol Cell Biol.* **11**: 1883-1893.
35. Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* **21**: 263-265.
36. de Bakker, P. I., R. Yelensky, I. Pe'er, S. B. Gabriel, M. J. Daly, and D. Altshuler. 2005. Efficiency and power in genetic association studies. *Nat Genet.* **37**: 1217-1223.
37. Falconer, D. S. 1989. Introduction to Quantitative Genetics. 3 ed. Longman Scientific and Technical, New York, NY.
38. Almasy, L. and J. Blangero. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* **62**: 1198-1211.
39. Whitlock, M. C. 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol.* **18**: 1368-1373.
40. Cox, N. J., M. Frigge, D. L. Nicolae, P. Concannon, C. L. Hanis, G. I. Bell, and A. Kong. 1999. Loci on chromosome 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nature Genetics.* **21**: 213-215.
41. Herbert, A., N. P. Gerry, M. B. McQueen, I. M. Heid, A. Pfeufer, T. Illig, H. E. Wichmann, T. Meitinger, D. Hunter, F. B. Hu, *et al.* 2006. A common genetic variant is associated with adult and childhood obesity. *Science.* **312**: 279-283.
42. Hotta, K., M. Nakamura, Y. Nakata, T. Matsuo, S. Kamohara, K. Kotani, R. Komatsu, N. Itoh, I. Mineo, J. Wada, *et al.* 2008. INSIG2 gene rs7566605 polymorphism is associated with severe obesity in Japanese. *J Hum Genet.* **53**: 857-862.

43. Oki, K., K. Yamane, N. Kamei, T. Asao, T. Awaya, and N. Kohno. 2008. The single nucleotide polymorphism upstream of insulin-induced gene 2 (INSIG2) is associated with the prevalence of hypercholesterolaemia, but not with obesity, in Japanese American women. *Br J Nutr.* 1-6.
44. Kim, S. K., L. Selleri, J. S. Lee, A. Y. Zhang, X. Gu, Y. Jacobs, and M. L. Cleary. 2002. Pbx1 inactivation disrupts pancreas development and in Ipfl-deficient mice promotes diabetes mellitus. *Nat Genet.* **30**: 430-435.
45. Miller, M., C. P. Cannon, S. A. Murphy, J. Qin, K. K. Ray, E. Braunwald. 2008. Impact of triglyceride levels beyond low-density lipoprotein cholesterol after acute coronary syndrome in the PROVE IT-TIMI 22 Trial. *J Am Col Cardiol.* **51**:724-730.
46. Kitsios, G.D. and E. Zintzaras. 2009. Genomic Convergence of Genome-wide Investigations for Complex Traits. *Ann Hum Genet.* 2009 Jul 9. [Epub ahead of print].

Table 1.

	MRC-OB Family Cohort		Replication Cohort	
	Average	Range	Average	Range
n	1560		920	
% female	71		80	
Age (years)	46	18-90	54	18-89
BMI (kg/m²)	31.9	17.1-75.3	32.6	15.9-86.1
Triglycerides (mg/dl)	128.0	27-457	135.8	36-424
LDL Cholesterol (mg/dl)	106.8	13-245	131.0	33-251

Table 2.

SNP	position	p-value MRC-OB family cohort	MAF	p-value Replication cohort	MAF
<i>rs6597489</i>	154696933	0.206	0.372		
<i>rs10901042</i>	154697319	0.184	0.371		
<i>rs9691590</i>	154701180	0.250	0.162		
rs9801154	154707375	0.126	0.028		
<i>rs9801301</i>	154709667	0.516	0.462		
<i>rs9690040</i>	154716649	0.045	0.372		
rs9691897	154718254	0.255	0.434		
SNP1	154718324	0.127	0.018		
<i>rs2721</i>	154718676	0.002	0.145	0.0008	0.151
rs1128640	154718751	0.542	0.114		
rs10271719	154718998	0.284	0.129		
rs12154226	154719026	0.146	0.425		
rs12154352	154719048	0.165	0.423		
rs12154437	154719051	0.153	0.423		
SNP2	154719102	0.146	0.427		
SNP3	154719128	0.119	0.422		
SNP4	154719186	0.111	0.423		
rs1128636	154719304	0.098	0.42		
rs9692271	154719881	0.185	0.423		
SNP5	154720345	0.464	0.078		
SNP6	154721022	0.464	0.001		
SNP7	154721030	0.207	0.001		
SNP8	154721574	0.202	0.133		
<i>rs9767875</i>	154722339	0.555	0.132		
<i>rs9768687</i>	154726371	0.148	0.117		
<i>rs9770068</i>	154730695	0.047	0.363		
<i>rs9769506</i>	154730788	0.037	0.421		
<i>rs12381375</i>	154736613	0.151	0.255		

Table 3.

rs7566605	rs2721	triglycerides (mg/dl)
CC and CG (n=763)	GG	105.6 ± 1.7
	GT	117.9 ± 1.8
	TT	138.4 ± 1.9*
GG (n=768)	GG	107.8 ± 1.7
	GT	108.9 ± 1.8
	TT	98.5 ± 1.6

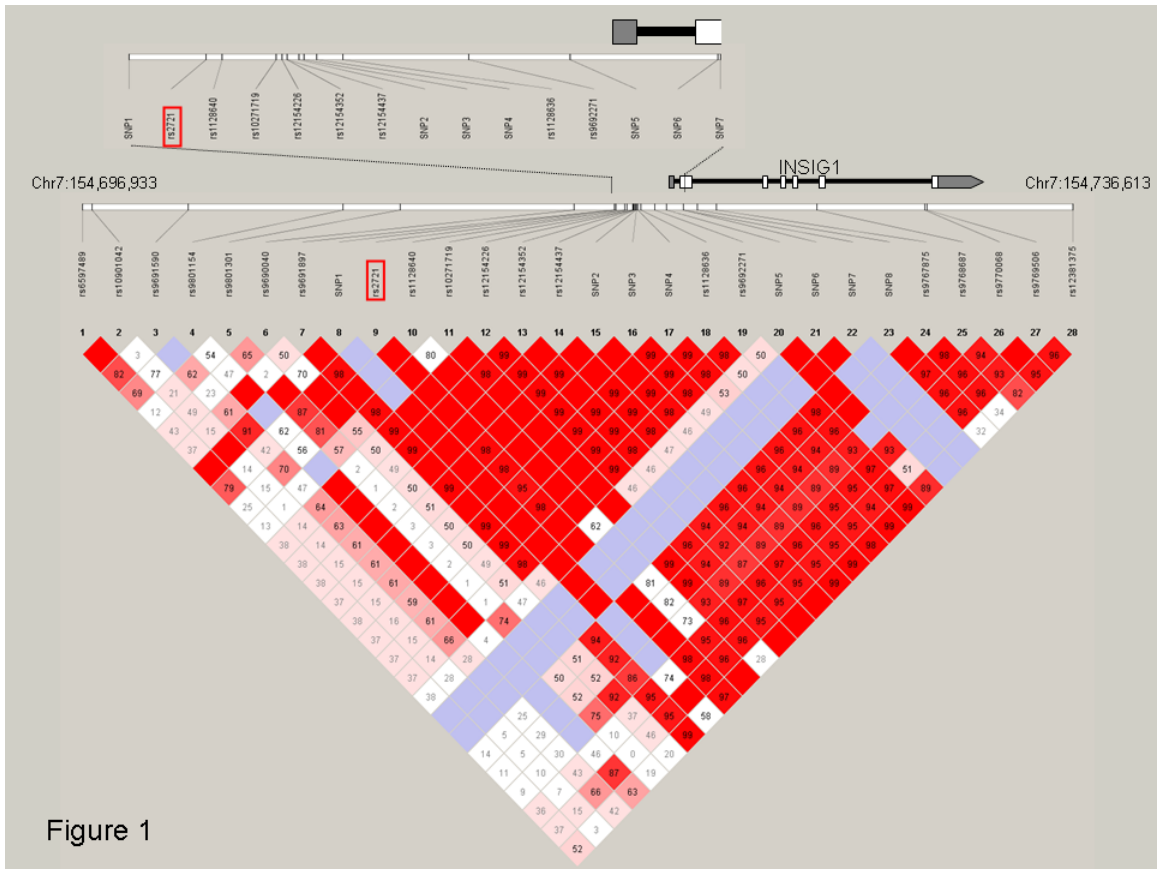


Figure 1

biotin	+	+	+	+	-	-
NE	+	+	-	-	+	+
allele	T	G	T	G	T	G



Figure 2.