

Familial combined hyperlipidemia is associated with upstream transcription factor 1 (*USF1*)

Päivi Pajukanta¹, Heidi E Lilja², Janet S Sinsheimer^{1,3}, Rita M Cantor^{1,4}, Aldons J Lusis^{1,5–7}, Massimiliano Gentile², Xiaoqun Joyce Duan⁸, Aino Soro-Paavonen⁹, Jussi Naukkarinen², Janna Saarela², Markku Laakso¹⁰, Christian Ehnholm², Marja-Riitta Taskinen⁹ & Leena Peltonen^{1,2}

Familial combined hyperlipidemia (FCHL), characterized by elevated levels of serum total cholesterol, triglycerides or both^{1,2}, is observed in about 20% of individuals with premature coronary heart disease¹. We previously identified a locus linked to FCHL on 1q21–q23 in Finnish families with the disease³. This region has also been linked to FCHL in families from other populations^{4–6} as well as to type 2 diabetes mellitus^{7–12}. These clinical entities have several overlapping phenotypic features, raising the possibility that the same gene may underlie the obtained linkage results. Here, we show that the human gene encoding thioredoxin interacting protein (*TXNIP*) on 1q, which underlies combined hyperlipidemia in mice¹³, is not associated with FCHL. We show that FCHL is linked and associated with the gene encoding upstream transcription factor 1 (*USF1*) in 60 extended families with FCHL, including 721 genotyped individuals ($P = 0.00002$), especially in males with high triglycerides ($P = 0.000009$). Expression profiles in fat biopsy samples from individuals with FCHL seemed to differ depending on their carrier status for the associated *USF1* haplotype. *USF1* encodes a transcription factor known to regulate several genes of glucose and lipid metabolism^{14–17}.

To identify the gene on 1q21 associated with FCHL, we initially sequenced four functionally relevant regional candidates: *TXNIP*, *USF1*, retinoid X receptor gamma (*RXRG*) and apolipoprotein A-II (*APOA2*). In parallel, we carried out a functionally unbiased genetic analysis of 60 single-nucleotide polymorphisms (SNPs) in 26 genes in 42 families with FCHL, including the 31 families in the original linkage study³. We then genotyped the ten SNPs most likely to be relevant in the extended sample of 60 families of FCHL (Supplementary Table 1 online). Fifty SNPs were located in a 5.8-Mb region flanking the peak markers *DIS104* and *DIS1677* (Fig. 1). All the families that we studied included a proband with severe

coronary heart disease and an abnormal lipid phenotype and an average of 5–6 members affected with FCHL.

We sequenced the entire *TXNIP* gene and the 2,000-bp upstream DNA region in 60 FCHL probands. Of the 20 SNPs identified, none resulted in amino acid changes, and all were rare, with a maximal 7% allele frequency. We also did not observe the nonsense mutation causing hyperlipidemia in mice¹³. We genotyped the four most common SNPs in the 60 families with FCHL but found no evidence of association

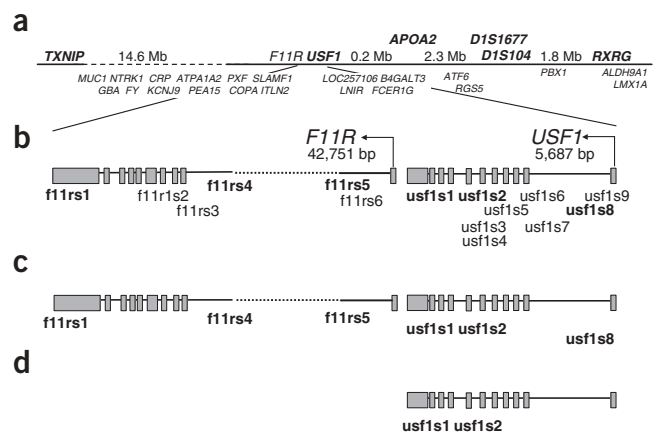


Figure 1 Schematic overview of the associated region on 1q21. (a) Genes and predicted genes (*LNIR* and *LOC257106*) in which we genotyped SNPs, as well as the locations of the peak linkage markers *DIS104* and *DIS1677* (ref. 3), are shown. The genes indicated in bold were also sequenced. (b) The SNPs genotyped for *F11R* and *USF1* (see Table 2 for distances, SNP numbers and LD clusters of these SNPs). (c) The SNPs associated with triglyceride levels in men and (d) the SNPs associated with FCHL and triglycerides in all family members.

¹Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, California 90095–7088, USA. ²Department of Molecular Medicine, National Public Health Institute, Finland and Department of Medical Genetics, University of Helsinki, Biomedicum, 00290 Helsinki, Finland. Departments of ³Biostatistics and Biostatistics, ⁴Pediatrics, ⁵Microbiology and Molecular Genetics, ⁶Medicine and ⁷Molecular Biology Institute; and ⁸UCLA-DOE Institute for Genomics and Proteomics, David Geffen School of Medicine at UCLA; University of California, Los Angeles, California 90095–7088, USA. ⁹Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland. ¹⁰University of Kuopio, Kuopio, Finland. Correspondence should be addressed to P.P. (ppajukanta@mednet.ucla.edu) or L.P. (leena.peltonen@ktl.fi).

Table 1 Multipoint HHRR and gamete competition analyses for the SNPs *usf1s1* (rs3737787) and *usf1s2* (rs2073658)

	FCHL all	Triglycerides all	FCHL men	Triglycerides men
Multi-HHRR	NS (NS)	0.05 (NS)	0.009 (NS)	0.00003 (0.003)
Gamete competition asymptotic <i>P</i> value	0.00002 (0.005)	0.00006 (0.008)	0.0004 (0.04)	0.0000009 (0.004)
Gamete competition (gene-dropping) empirical <i>P</i> value	0.00004	0.00006	0.0004	0.00001

All *P* values are for simultaneous analysis of both SNPs. NS, nonsignificant. The first *P* values were obtained from analysis of 60 extended families with FCHL (721 genotyped subjects), and the values given in parentheses from analysis of 42 nuclear families with FCHL (238 genotyped subjects). Gene-dropping analysis was done for only the 60 extended families with FCHL using at least 50,000 simulations. The segregating haplotype was 1-1 (1 indicates the common allele) in all gamete competition analyses above.

between individual SNPs or their haplotypes and FCHL or triglycerides ($P > 0.1$; **Supplementary Table 2** online). For *RXR*G, sequencing identified 11 SNPs, none of which resulted in missense or nonsense variants. Two informative SNPs, rs157870 in intron 2 and rs2134095 in exon 8, were not associated with FCHL or triglycerides in 60 families (**Supplementary Table 1** online). Sequencing of *APOA2* identified six noncoding variants, and we genotyped four of these in 42 families but observed no association (**Supplementary Table 1** online). Because a microsatellite within *APOA2* produced a linkage signal in our original report³, however, we genotyped two *APOA2* SNPs in all 60 families with FCHL. No evidence emerged for association between any individual SNPs or their haplotypes and FCHL or triglycerides traits.

Sequencing of *USF1* identified 23 SNPs (**Supplementary Table 3** online), none resulting in amino acid changes. Initially, we genotyped three SNPs in 42 families with FCHL: *usf1s1* (exon 11), *usf1s2* (intron 7) and *usf1s7* (exon 2; for SNP numbers, see **Tables 1** and **2**). Both *usf1s1* and *usf1s2* showed evidence for linkage (lod scores of 3.5 and 2.0, respectively, for FCHL and 3.7 and 2.0, respectively, for triglycerides), and combined analysis of the two SNPs showed some evidence for association with FCHL ($P = 0.005$) and triglycerides ($P = 0.008$) using the gamete competition test¹⁸ (**Table 1**). *usf1s1* and *usf1s2* also suggested an association in men with high triglycerides; the individual and combined analyses of these SNPs gave *P* values between 0.02 and 0.003 (**Tables 1** and **2**).

To further investigate the region, we genotyped a total of 15 SNPs (4 of which were also genotyped in 60 extended families based on

their evidence of linkage or their pattern of linkage disequilibrium (LD) **Supplementary Table 3** online and **Table 2**). Again, combined analysis of *usf1s1* and *usf1s2* showed substantial evidence for linkage and association with both FCHL ($P = 0.00002$) and triglycerides ($P = 0.00006$; **Table 1**). In all affected family members, the association with both FCHL and triglycerides was limited to *usf1s1* and *usf1s2*, 1,239 bp apart (**Table 1**), with the remaining SNPs providing no evidence for association. For both FCHL and triglycerides, the trait-segregating haplotype included the common alleles (1-1) of *usf1s1* and *usf1s2*. For association with triglycerides in men, *usf1s1* and *usf1s2* yielded *P* values of 0.002–0.00001 (**Table 2**), and the combined analyses of the two SNPs yielded *P* values of 0.00003–0.0000009 (**Table 1**). Four other SNPs, three of them in strong LD with *usf1s1* and *usf1s2* ($P < 0.00002$), showed evidence of association with triglycerides in men, extending the associated region to 46 kb, including the adjacent gene F11 receptor (*F11R*; **Table 2**). No association was obtained with the tested SNPs residing outside the *F11R-USF1* region (**Supplementary Table 1** online).

To address the allelic diversity in this critical region, we tested haplotypes of several SNPs for association using a variety of test statistics, including the HBAT-e (ref. 19), genotype-PDT²⁰ and multipoint haplotype-based haplotype relative risk (HHRR)²¹ tests, and obtained consistent evidence for association and shared haplotypes with *usf1s1* and *usf1s2* (**Table 3**). Transmission of the haplotype of the rare alleles (2-2) to the affected subjects was reduced ($P = 0.004$), suggestive of a protective role for this haplotype (**Table 3**).

Table 2 Association analyses of individual SNPs in the *F11R-USF1* region with triglycerides and FCHL in men

SNP	rs number	Distance (bp)	Heterozygosity/rare allele		Triglycerides HHRR	Triglycerides gamete	FCHL HHRR	FCHL gamete	LD cluster (I–V)
			frequency in all family members						
f11rs1	rs836	1,361	0.41/0.28		0.03	0.009	NS	0.03	I
f11rs2	rs790056	1,561	0.36/0.24		NS	0.03	NS	NS	II
f11rs3	rs790055	25,608	0.35/0.23		NS	NS	NS	NS	II
f11rs4	hCV1459766	10,572	0.38/0.26		0.06	0.04	NS	NS	I
f11rs5	rs4339888	1,246	0.43/0.31		0.02	0.003	NS	0.09	I
f11rs6	rs3766383	951	0.25/0.15		NS	NS	NS	NS	III
usf1s1	rs3737787	1,239	0.45/0.34		0.0009 (0.01)	0.00001 (0.008)	0.04 (NS)	0.05 (NS)	I
usf1s2	rs2073658	12	0.44/0.33		0.002 (0.02)	0.00006 (0.02)	0.04 (NS)	NS (NS)	I
usf1s3	rs2516841	17	0.40/0.28		NS	NS	NS	NS	II
usf1s4	rs2073657	526	0.48/0.41		NS	NS	NS	NS	IV
usf1s5	rs2516840	1,443	0.41/0.29		NS	NS	NS	NS	II
usf1s6	rs2073653	361	0.25/0.14		NS	0.08	NS	NS	III
usf1s7	rs2516839	1,249	0.47/0.39		NS (NS)	0.04 (NS)	NS (NS)	NS (NS)	IV
usf1s8	rs2516838	279	0.40/0.28		0.01 (0.05)	0.05 (0.03)	NS (NS)	NS (NS)	V
usf1s9	rs1556259		0.23/0.13		NS	NS	NS	NS	III

All results represent *P* values. NS, nonsignificant; gamete, gamete competition test. LD cluster number in the last column indicates the clusters of SNPs showing strong intermarker LD ($P \leq 0.00002$) in the male probands with high triglycerides (>90th age–sex specific population percentile); the SNPs carrying the same cluster number are in strong pairwise LD. SNPs indicated in bold were genotyped in the 60 extended families with FCHL, and the values in parentheses were obtained from analysis of the 42 nuclear families with FCHL. All other results were obtained in the 42 nuclear FCHL families.

Table 3 Haplotype analyses in men with elevated triglyceride levels using HBAT

Test	Haplotype of SNPs f11rs4–usf1s2	Haplotype of SNPs usf1s1–usf1s2	Haplotype of SNPs usf1s1–usf1s5
HBAT–o	0.03 (1–1–1–1)	0.0007 (1–1) 0.004 (2–2 (protective), significantly less transmitted to the affected individuals)	NS (0.07) (1–1–1–1–1)
HBAT–e	0.009 (1–1–1–1)	0.02 (1–1)	NS (0.2) (1–1–1–1–1)
Multilocus geno-PDT	0.02	0.002	NS (0.7)
Multi-HHRR	0.0002	0.00003	0.04

The inter-SNP distances and corresponding *r*s numbers for the SNPs f11rs4–f11rs6 and usf1s1–usf1s5 are shown in **Table 2**. 1, common allele; NS, nonsignificant. The *P* values from the HBAT program indicate the probability that the particular haplotype is transmitted to the affected individuals using the option –o (optimize offset) or option –e (empirical test). Multilocus geno-PDT indicates a genotype-based association test for general pedigrees (genotype Pedigree Disequilibrium Test). The multipoint HHRR analysis tests the hypothesis of homogeneity of marker allele distributions between transmitted and nontransmitted alleles of the SNPs.

To address whether *USF1* was associated with triglyceride levels rather than with the complex FCHL phenotype, we tested the association of the usf1s1–usf1s2 combination with three qualitative lipid traits: increased apolipoprotein B (apoB), increased total cholesterol and small low-density lipoprotein (LDL) peak particle size. For apoB, we obtained *P* values of 0.00003 or 0.0007 for association between all individuals or affected men, respectively, and the susceptibility haplotype of common alleles, using the gamete competition analysis. For total cholesterol, the corresponding *P* values were 0.0001 and 0.007, respectively, and for LDL peak particle size, 0.002 and 0.01, respectively. These results imply that *USF1* contributes to the complex abnormal lipid phenotype characteristic of FCHL.

To evaluate the significance of the gamete competition results, we calculated empiric *P* values for all analyses involving multiple SNPs (**Table 1**) using gene dropping. These *P* values agreed well with the asymptotic *P* values of the gamete competition analyses (**Table 1**), indicating that the results were not artifacts of asymptotic approximations with sparse data.

We next compared expression profiles in fat biopsy samples from six individuals with FCHL carrying the *USF1* risk haplotype 1–1 with those from four individuals with FCHL who were homozygous with respect to the putative protective haplotype 2–2 using the Affymetrix HGU133A array. Applying highly stringent criteria, we determined that 25 genes were upregulated and 73 genes were downregulated in individuals carrying the risk haplotype. We detected no haplotype-dependent differences in *TXNIP* expression. To lend biological relevance to these findings, we examined the lists of differentially expressed genes for over-representation of any functional classes. Only three classes were significantly over-represented among the upregulated genes in carriers of the risk haplotype. These were primarily genes involved in fat metabolism. We observed a prominent downregulation of immune-response genes (**Fig. 2** and **Supplementary Table 4** online). These data suggest that the *USF1* risk haplotype affects the expression profiles in fat biopsy samples. We verified the expression of *USF1* in fat biopsy samples using quantitative real-time PCR, but we observed no differences in the relative expression levels of *USF1* between individuals with FCHL carrying the risk haplotype and those with FCHL not carrying the risk haplotype (data not shown).

Because we could not establish an obvious functional change due to the FCHL-associated *USF1* allele, we investigated the genomic sequence flanking the risk haplotype for potential functional domains and identified a 60-bp sequence element present in 91 human genes (**Fig. 3a** and **Supplementary Table 5** online). This 60-bp region is highly conserved and is found in pufferfish and in *Caenorhabditis elegans* but not in *Drosophila melanogaster* or in *Saccharomyces cerevisiae*. Analysis of domain annotation indicated an enrichment of domains involved in protein modification (*n* = 16 genes) and nucleic acids (*n* = 35). Accordingly, annotations about the biological process implied an involvement in nucleic acid metabolism (*n* = 18), as well as transcription and signal transduction (*n* = 33). The SEAP reporter assay indicated that this 60-bp element has an effect on transcription *in vitro* in the forward orientation. The reverse orientation resulted in a transcription efficiency comparable to that of the negative control (**Fig. 3b**), suggestive of a *cis*-acting regulator rather than a direction-independent enhancer element. These data are suggestive of a putative regulatory element in the immediate vicinity of the *USF1* risk haplotype.

We identified the gene associated with FCHL underlying the linkage signal on 1q21 (ref. 3–6). We excluded *TXNIP*, a causative gene for combined hyperlipidemia in mice¹³, and found that *USF1* had significant association, linkage and shared haplotypes of the disease-associated alleles. The strongest evidence for association was

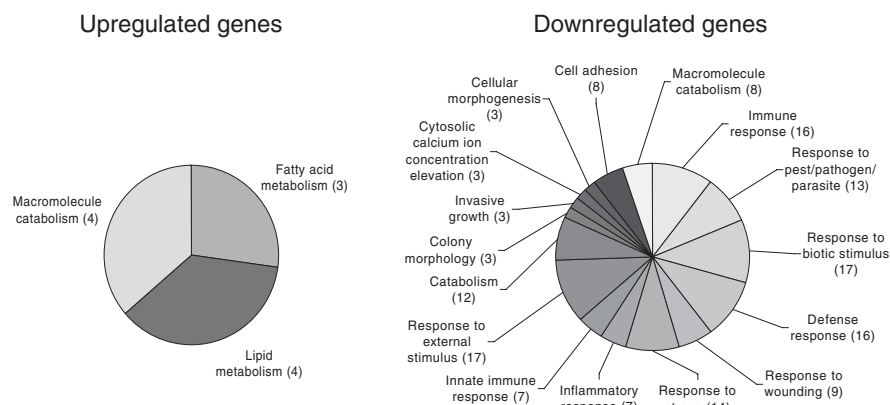


Figure 2 Distribution of genes according to functional category for the 16 upregulated and 60 downregulated genes for which annotation information for the Gene Ontology²⁶ class Biological process was available. In this analysis, lists of differentially expressed genes were examined for over-representation of functional classes, as defined by the Gene Ontology consortium, using the EASE tool²⁷. Only categories scoring a statistically significant EASE score (<0.05) for over-representation are shown. Complete results of the EASE analysis including the corresponding EASE scores (*P* values) and the lists of genes in every significant category are given in **Supplementary Table 4** online.

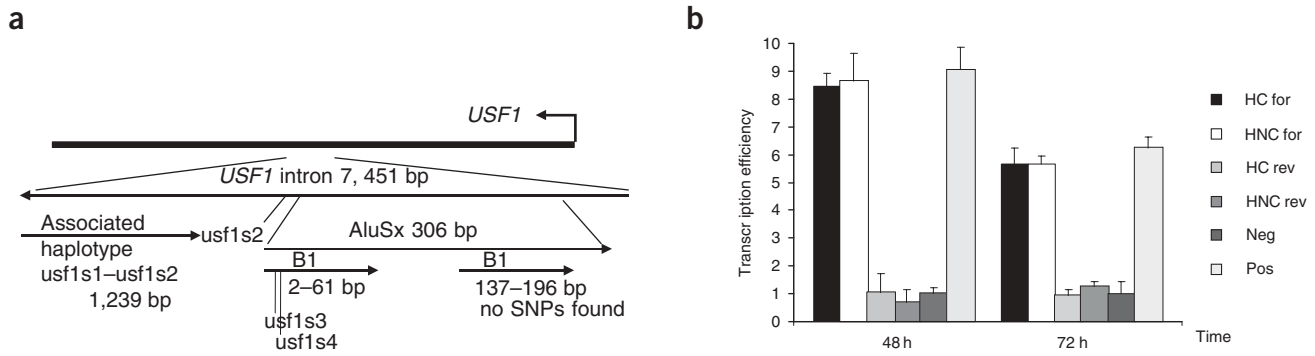


Figure 3 Investigation of the genomic sequence flanking the risk haplotype. **(a)** Identification of a 60-bp conserved sequence element in intron 7 of *USF1*. The SNP *usf1s2*, forming part of the risk haplotype, resides adjacent (8 bp) to a 306-bp *AluSx* repeat. Two parts (2–61 bp and 137–196 bp) of this *AluSx* repeat show sequence similarity with the mouse B1 repeat. When compared by BLAST against the mouse sequence databases, these two parts of the *AluSx* sequence identify numerous mouse expressed-sequence tags, due to the B1 element located in the untranslated region of the mouse mRNA. A total of 91 human genes, including *USF1*, have this 60-bp part of *AluSx* located either on the coding strand (43 genes) or on the opposite strand (48 genes). A complete list of the 91 human genes and their individual *P* values are given in **Supplementary Table 5** online. **(b)** Transcription efficiency of a 268-bp region in intron 7 of *USF1* containing the 60-bp conserved sequence and the *usf1s2* SNP. DNAs from one homozygous carrier of the susceptibility haplotype (1–1; HC) and one homozygous noncarrier (2–2; HNC) were cloned to the SEAP reporter system in both forward (for) and reverse (rev) orientations. Culture medium from cells transfected with the pSEAP2-Basic vector was used as a negative control (Neg), and culture medium from cells transfected with the pSEAP2-Control vector was used as a positive control (Pos). The SEAP protein was monitored 48 and 72 h after transfection. Error bars represent s.d. of one experiment done in triplicate. The size of the bar indicates the increase in transcriptional activity when compared with the negative control, which is set to 1.

seen in males with elevated levels of triglycerides. Maleness is a known risk factor for coronary heart disease, and recent studies also suggest gender-specific differences in the dyslipidemic phenotypes of FCHL²². Our preliminary functional data support these genetic data, as the *USF1* risk haplotype seems to affect the expression profiles in fat biopsy samples from individuals with FCHL. We also identified a new putative regulatory element in *USF1* flanking the susceptibility haplotype. We observed no differences in steady-state *USF1* expression levels between individuals with FCHL carrying the *USF1* risk haplotype and those not carrying the risk haplotype, suggesting that the associated allele does not have a direct effect on *USF1* transcription in adipose tissue. Given the complexity of transcriptional regulation, the potential effect of the *USF1* risk allele probably results from tissue- or cell type-specific differences due to special local stimuli, which may not be accessible in fat biopsy samples. Additional studies are warranted to address the functional differences between different *USF1* alleles and their relevance to the FCHL phenotype.

Our study included 60 extended families, who fulfilled the strictest phenotypic criteria for FCHL¹ and originated from a relatively isolated population²³ with an increased probability of extended LD, making the family-based association analysis we used suitable. The SNP analysis of alleles associated with FCHL would restrict the critical region to 1,239 bp in *USF1*, whereas the association in males with elevated triglycerides extends to the adjacent gene *F11R*. The known functions of *F11R*, associated with T-cell migration²⁴, make it a less likely candidate for the gene underlying FCHL. We cannot yet positively confirm a single associated causative variant, but we identified several associated SNPs in tight LD and a common SNP haplotype defining the disease-associated *USF1* allele in the Finnish families with FCHL.

USF1 belongs to the basic helix-loop-helix leucine zipper family, interacts with its target DNA as a homodimer or heterodimer with *USF2* and recognizes a CACGTG motif called E box in the promoter of the target genes, resulting in transcriptional activation in response to various stimuli, such as glucose and dietary carbohydrates^{14,15}. Target

genes of *USF1* include several apolipoproteins (CIII, AII and E), hormone-sensitive lipase, fatty acid synthase, glucokinase, the glucagon receptor, ATP-binding cassette sub-family A (ABC1) member 1, and renin, making *USF1* a good candidate for involvement with the central clinical features of FCHL and type 2 diabetes mellitus: glucose intolerance, hyperlipidemia, insulin resistance and hypertension. The concept that *USF1* affects the complex lipid phenotype of FCHL, and not only one lipid trait, is supported by our findings of allelic associations of the *usf1s1-usf1s2* risk haplotype with triglycerides, apoB, total cholesterol and LDL peak particle size.

METHODS

Subjects. The Finnish families with FCHL were recruited in the Helsinki, Turku and Kuopio University Central Hospitals³. Each subject provided written informed consent. We collected all samples in accordance with the Helsinki declaration, and the ethics committees of the participating centers approved the study design. We described the inclusion and exclusion criteria for the FCHL probands in detail earlier³. For the FCHL trait, we scored family members as affected according to the same diagnostic criteria that we used in our original linkage study³ using the Finnish age- and sex-specific 90th percentiles for high total cholesterol and high triglycerides, available at our website. These ascertainment criteria are fully comparable with the original criteria¹. For analysis of triglycerides, we coded family members with triglyceride levels \geq the 90th percentile of the Finnish age- and sex-specific population as affected. We also tested the allelic association of the *usf1s1-usf1s2* SNPs using apoB, LDL peak particle size and total cholesterol traits. For apoB and total cholesterol, we used the 90th age- and sex-specific Finnish population percentiles. For LDL peak particle size, we considered individuals with particle size ≤ 25.5 nm to be affected.

Biochemical analyses. We measured serum lipid parameters and LDL peak particle size as described earlier^{3,25}. In the 60 families with FCHL, DNA and lipid measurements were available for 721 and 771 family members, respectively. Using the 90th Finnish age- and sex-specific population percentiles, 226 family members had elevated total cholesterol levels, 220 had elevated triglyceride levels, 321 had elevated total cholesterol and/or triglyceride levels and 125 had elevated total cholesterol and triglyceride levels. A total of 96 men and 124 women had elevated triglyceride levels.

Sequencing, genotyping and sequence annotations. We sequenced all 60 FCHL probands for *TXNIP* and the 31 probands of the original linkage study³ for *APOA2*, *RXRG* and *USF1*. For *TXNIP* and *USF1*, we also sequenced 2,000 bp upstream from the 5' end of the gene. For *USF1*, we also sequenced the DNA binding domain in the remaining 29 probands. For all genes, we sequenced both exons and introns, except for the large 44,261-bp gene *RXRG*, in which we sequenced all exons and 100-bp exon-intron boundaries. Sequencing was done using the automated DNA sequencer ABI 377XL (Applied Biosystems). We assembled the sequence contigs using the Sequencher software (GeneCodes). We used the dbSNP and CELERA databases and sequencing to select SNPs for genotyping with pyrosequencing or solid-phase minisequencing³. We determined the physical order of the markers and genes using the University of California Santa Cruz Genome Browser. We will submit new SNPs to public databases (National Center for Biotechnology Information). We tested all SNPs for possible violation of Hardy-Weinberg equilibrium in three groups (all family members, probands, and spouses). We downloaded annotation data of the *Alu* elements from the University of California Santa Cruz Genome Browser, which uses RepeatMasker to screen DNA sequences for interspersed repeats. We determined the positions of the 60-bp sequence on these *Alu* elements using BLAST. We downloaded other annotation data from LocusLink.

Expression array analysis of adipose tissue. We selected six individuals with FCHL carrying the susceptibility haplotype and four individuals with FCHL who were homozygous with respect to the protective haplotype for assessment of gene expression. The six carriers of the susceptibility haplotype were from six individual families. The four homozygous individuals were two sibling pairs from two families. We collected biopsy samples from umbilical subcutaneous adipose tissue under local anesthesia to collect 50–2,000 mg of adipose tissue. We extracted RNA using STAT RNA-60 reagent (Tel-Test), treated it with DNase I and purified it with RNeasy Mini Kit columns (Qiagen). We assessed the quality of the RNA using the RNA 6000 Nano assay in the Bioanalyzer (Agilent) monitoring for ribosomal S28/S18 RNA ratio and signs of degradation. We measured the concentration and the A_{260}/A_{280} ratio of the samples using a spectrophotometer, the acceptable ratio being 1.8–2.2. We carried out RNA labeling, array experiments with Human genome U133 arrays, scanning and primary data analysis using the standard protocol by Affymetrix with minor modifications.

We analyzed scanned images with Affymetrix Microarray Suite 5 (Affymetrix) software using the Statistical Expression Algorithm. We applied global scaling to a target intensity of 100 to all arrays and then processed data with GeneSpring 5.0 data analysis software (Silicon Genetics). For each probe array, we applied a per-gene normalization so that signal intensities were divided by the median intensity calculated using all ten probe arrays. We determined cut-off values to discriminate low-quality data separately for each haplotype group by dividing the base value with the proportional value estimated using the Cross Gene Error Model implemented in GeneSpring. To identify differentially expressed genes between the two haplotypes, we calculated ratios of averaged normalized intensities. Differences were considered as significant if the resulting ratio fell at least three standard deviations outside the average ratio calculated from the distribution of the \log_{10} of the ratios. Only genes scored as present in all ten samples, or as absent or marginal in all cases and present in all the controls (or vice versa), were included. We retrieved annotation information defining the biological processes for individual genes from the classifications provided by the Gene Ontology consortium²⁶. We statistically evaluated the enrichment of categories represented in each gene list, compared with the proportion observed in the total population of genes on the probe array, using the Expression Analysis Systematic Explorer (EASE) tool²⁷, with the threshold value set to 3. We calculated the test statistic using Fisher's exact test. To maximize robustness, we calculated an EASE score (P value) where the Fisher exact probabilities were adjusted so that categories supported by few genes were strongly penalized and categories supported by many genes were negligibly penalized. EASE scores (P values) below 0.05 were considered statistically significant.

Quantitative real-time PCR analysis of *USF1*. We selected two individuals with FCHL carrying the susceptibility haplotype and two individuals with FCHL without the haplotype for assessment of *USF1* expression in adipose

tissue using the SYBR-Green assay (Applied Biosystems). We carried out two-step RT-PCR using TaqMan Gold RT-PCR kit. Primer sequences are available on request. We carried out the reactions in triplicate using the ABI Prism 7900 HT Sequence Detection System and analyzed the data using Sequence Detector version 2.0 software.

Reporter gene analysis for the transcription efficiency. We analyzed the effect of the 60-bp conserved sequence on transcription *in vitro* using the SEAP reporter system (Clontech) in COS cells. We cloned the target sequences into the pSEAP2-Enhancer vector. We verified the correct allele and orientation in each construct by sequencing. We monitored the SEAP protein using the fluorescent substrate 4-methylumbelliferyl phosphate in a fluorescent assay according to the manufacturer's instructions. Data are representative of at least two independent experiments.

Statistical analyses. We carried out parametric and nonparametric linkage analyses using the MLINK program of the LINKAGE package²⁸ and the SIB-PAIR program as implemented by the ANALYZE package²⁹ using the same parameters as earlier³. For each marker, we estimated allele frequencies from all individuals using the DOWNFREQ program³⁰.

We tested the SNPs for association using the HHRR²¹ test and the gamete competition test¹⁸. To minimize the number of tests needed, we tested SNPs outside the *USF1-F11R* region using only the HHRR test when analyzing males with elevated triglycerides and with FCHL. The HHRR analysis, done with the HRR LAMB program, tests the homogeneity of marker allele distributions between transmitted and nontransmitted alleles. The multipoint HHRR analysis, done with the HRRMULT program, tests the same hypothesis using several SNPs. The gamete competition test is a generalization of the transmission disequilibrium test and views transmission of marker alleles to affected children as a contest between the alleles, making effective use of full pedigree data. The gamete competition method is not purely a test of association, because the null hypothesis is no association and no linkage, and thus linkage in itself also affects the observed P value. P values based on asymptotic approximations can be biased when data used to calculate them are relatively sparse. To confirm that the gamete competition results were significant, we also calculated empirical P values for all analyses involving multiple SNPs (Table 1) using gene-dropping. In gene-dropping, the founder genotypes are assigned using the estimated allele frequencies assuming Hardy-Weinberg equilibrium and linkage equilibrium. The offspring genotypes are assigned assuming mendelian segregation. Thus, gene-dropping is done under the null hypothesis of linkage equilibrium and no linkage. To calculate an empirical P value, gene-dropping is carried out multiple times. Here, at least 50,000 simulations were carried out for each analysis. The likelihood ratio test statistic (LRT) from each gene-dropping iteration is compared to the LRT for the observed data. The empirical P value is the proportion of iterations in which the gene-dropping LRT equaled or exceeded the observed LRT. In general, the obtained empirical P values of gene-dropping are more conservative than asymptotic P values for small sample sizes.

We used the HBAT¹⁹ program, with the options optimize offset ($-o$) and empirical test ($-e$), to test for association between haplotypes and the trait. The option $-o$ measures not only preferential transmission of the susceptibility haplotype to affected individuals but also less preferential transmissions to unaffected individuals, making it useful here because, in these extended families, the unaffected individuals also provide important information. The $-e$ option leads to a test of association given linkage and thus gives an empirical estimation of the variance. This test statistic takes the linkage information into account. We also carried out the genotype Pedigree Disequilibrium Test²⁰, which provides a genotype-based association test for general pedigrees, for a combination of genotypes from selected *USF1* SNPs (Table 3). We tested LD between the marker genotypes for SNPs in the *F11R-USF1* region using the Genepop v3.1b program, option 2, at their website.

URLs. Supplementary Tables 1–5 are also available at our websites (<http://www.genetics.ucla.edu/labs/pajukanta/fchl/chr1/> and <http://www.ktl.fi/mol/wwpub.htm>). The raw data for the complete set of probe arrays can be accessed through the Gene Expression Omnibus at National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/geo>). The Finnish 90th age- and sex-specific percentile values for total cholesterol, triglycerides

and apoB are available at the website of the National Public Health Institute of Finland (<http://www.ktl.fi/molbio/wwwpub/fchl/genomescan>). We used Genepop (<http://wbiomed.curtin.edu.au/genepop/index.html>) to calculate intermarker LD.

GEO accession number. GSE590.

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

ACKNOWLEDGMENTS

We thank the families for their participation in this study; K. Porkka, J. Viikari, K. Ylitalo, J. Vakkilainen, I. Nuotio and J. Pihlajamäki for their participation in the initial study organization and family collection; A. Jokiahio, E. Korshennikova, P. Tainola, T. Airaksinen, L. Puhakka, T. Manninen, A. Kuraishy, J. Schmidt, N. Aguda and A. Korhonen for technical assistance; J. Wolford for SNP information; S. Horvath for his help with the HBAT program; and M. Perola, T. Hiekkalinna, L. Quiogue and J. Saharinen for their help with bioinformatics. This study was funded by the US National Institute of Health (NHLBI), the Laubisch Foundation, the Center of Excellence of Disease Genetics of the Academy of Finland, the Maud Kuistila Foundation, the Finnish Cultural Foundation, the Helsinki University Central Hospital Research Foundation, the Finnish Heart Foundation and the Duodecim Foundation. X.J.D. is supported by Alfred P. Sloan Foundation/Department of Energy Postdoctoral Fellowship. L.P. is Gordon and Virginia MacDonald Distinguished Chair in Human Genetics at UCLA.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 17 August 2003; accepted 27 January 2004

Published online at <http://www.nature.com/naturegenetics/>

- Goldstein, J.L., Schrott, H.G., Hazzard, W.R., Bierman, E.L. & Motulsky, A.G. Hyperlipidemia in coronary heart disease II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J. Clin. Invest.* **52**, 1544–1568 (1973).
- Nikkilä, E.A. & Aro, A. Family study of serum lipids and lipoproteins in coronary heart disease. *Lancet* **1**, 954–959 (1973).
- Pajukanta, P. *et al.* Linkage of familial combined hyperlipidemia to chromosome 1q21q23. *Nat. Genet.* **18**, 369–373 (1998).
- Coon, H. *et al.* Replication of linkage of familial combined hyperlipidemia to chromosome 1q with additional heterogeneous effect of apolipoprotein A-I/C-III/A-IV locus: the NHLBI family heart study. *Arterioscler. Thromb. Vasc. Biol.* **20**, 2275–2280 (2000).
- Pei, W. *et al.* Support for linkage of familial combined hyperlipidemia to chromosome 1q21–q23 in Chinese and German families. *Clin. Genet.* **57**, 29–34 (2000).
- Allayee, A. *et al.* Locus for elevated apolipoprotein B levels on Chromosome 1p31 in families with familial combined hyperlipidemia. *Circ. Res.* **90**, 926–931 (2002).
- Elbein, S.C., Hoffman, M.D., Teng, K., Leppert, M.F. & Hasstedt, S.J. A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes* **48**, 1175–1182 (1999).
- Hanson, R.L. *et al.* An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am. J. Hum. Genet.* **63A**, 1130–1138 (1998).
- Vionnet, N. *et al.* Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q-qter and independent replication of a type 2-diabetes locus on chromosome 1q21–q24. *Am. J. Hum. Genet.* **67**, 1470–1480 (2000).
- Wiltshire, S. *et al.* A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am. J. Hum. Genet.* **69**, 553–569 (2001).
- Hsueh, W.C. *et al.* Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21–q24. *Diabetes* **52**, 550–507 (2003).
- Watanabe, R.M. *et al.* The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. *Am. J. Hum. Genet.* **67**, 1186–1200 (2000).
- Bodnar, J.S. *et al.* Positional cloning of the combined hyperlipidemia gene *Hyp1p1*. *Nat. Genet.* **30**, 110–116 (2002).
- Casado, M., Vallet, V.S., Kahn, A. & Vaulont, S. Essential role in vivo of upstream stimulatory factors for a normal dietary response of the fatty acid synthase gene in the liver. *J. Biol. Chem.* **274**, 2009–2013 (1999).
- Vallet, V.S. *et al.* Differential roles of upstream stimulatory factors 1 and 2 in the transcriptional response of liver genes to glucose. *J. Biol. Chem.* **274**, 20175–20179 (1998).
- Ribeiro, A., Pastier, D., Kardassis, D., Chambaz, J. & Cardot, P. Cooperative binding of upstream stimulatory factor and hepatic nuclear factor 4 drives the transcription of the human apolipoprotein A-II gene. *J. Biol. Chem.* **274**, 1216–1225 (1999).
- lynedjian, P.B. Identification of upstream stimulatory factor as transcriptional activator of the liver promoter of the glucokinase gene. *Biochem. J.* **333**, 705–712 (1998).
- Sinsheimer, J.S., Blangero, J. & Lange, K. Gamete competition models. *Am. J. Hum. Genet.* **66**, 1168–1172 (2000).
- Laird, N., Horvath, S. & Xu, X. Implementing a unified approach to family based tests of association. *Genet. Epidemiol.* **19**, 36–42 (2000).
- Martin, E.R., Bass, M.P., Gilbert, J.R., Pericak-Vance, M.A., & Hauser ER. Genotype-based association test for general pedigrees: the genotype-PDT. *Genet. Epidemiol.* **25**, 203–213 (2003).
- Terwilliger, J.D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346 (1992).
- Soro, A., Jauhainen, M., Ehnholm, C. & Taskinen, M.R. Determinants of low HDL levels in familial combined hyperlipidemia. *J. Lipid Res.* **44**, 1536–1544 (2003).
- Peltonen, L., Pekkarinen, P. & Aaltonen, J. Messages from an isolate: lessons from the Finnish gene pool. *Biol. Chem.* **376**, 697–704 (1995).
- Ostermann, G., Weber, K.S., Zerneck, A., Schroder, A. & Weber, C. JAM-1 is a ligand of the beta(2) integrin LFA-1 involved in transendothelial migration of leukocytes. *Nat. Immunol.* **3**, 151–158 (2002).
- Vakkilainen, J. *et al.* LDL particle size in familial combined hyperlipidemia: effects of serum lipids, lipoprotein-modifying enzymes, and lipid transfer proteins. *J. Lipid Res.* **43**, 598–603 (2002).
- The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
- Hosack, D.A., Dennis, G. Jr., Sherman, B.T., Lane, H.C. & Lempicki, R.A. Identifying biological themes within lists of genes with EASE. *Genome Biol.* **4**, R70.1–R70.8 (2003).
- Lathrop, G.M., Lalouel, J.-M., Julier, C.A. & Ott, J. Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81**, 3443–3446 (1984).
- Göring, H.H. & Terwilliger, J.D. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum. Biol.* **72**, 63–132 (2000).
- Göring, H.H. & Terwilliger, J.D. Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *Am. J. Hum. Genet.* **66**, 1298–1309 (2000).