Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Ober C, Tan Z, Sun Y, et al. Effect of variation in *CHI3L1* on serum YKL-40 level, risk of asthma, and lung function. N Engl J Med 2008;358:1682-91.

Supplementary Materials

Genotyping and QC methods.

The Hutterites in this study were genotyped with the Affymetrix GeneChip® 500k Mapping Array, using both the early access and commercial Affymetrix GeneChip® 500k Mapping Array at The University of Chicago. A set of 421,374 autosomal SNPs were present on both sets of chips. Another 1,423 nsSNPs were genotyped at the NHLBI Resequencing and Genotyping Service (Johns Hopkins University) using a custom 1,536 SNP oligo pool and BeadArray method, as previously described[1]. In the combined set of SNPs, 131,049 were not further studied because either they were monomorphic (N=52,732) or had MAFs <5% (N=58,152) in the Hutterites. The remaining 310,490 SNPs were subjected to quality control checks. An additional 20,165 SNPs were excluded because either they had call rates <90% (N=3,614), they deviated from Hardy-Weinberg expectations at $p<0.001$ (correcting for the Hutterite inbreeding and population structure) (N=5,082), or because they generated ≥5 Mendelian errors (N=11,469), yielding a set of 290,325 markers with a median inter-maker spacing of 4.3 kb (range 17 bp – 22.97 kb).

Association testing in the Hutterites.

The natural log of serum YKL-40 level was used for the heritability and association studies; age and sex were included as covariates in all analyses.[2] The heritability of serum YKL-40 was estimated using a variance component maximum likelihood method.[3] At each SNP, we used the general two-allele model test of association in the entire pedigree, keeping all inbreeding loops

intact, as described.[4] SNP-specific p-values were determined based on Gaussian theory;[4] genome-wide p-values were determined by a Monte Carlo permutation-based test that preserves the covariance structure due to relatedness of individuals and assesses significance in the presence of multiple, dependent tests while guarding against deviations from normality in the data. We used 100 permutations to generate the empirical distribution of p-values and considered a p-value to be genome-wide significant if it was equal to or smaller than the 5% quantile of the permutation-based empirical distribution of the global minimum p-value.

Table S1. Mean ln serum YKL-40 levels (ng/ml) in COAST children by age and *CHI3L1* -131C→G genotype. N, sample size; SE, standard error. See Figure 3 in manuscript.

	CC Genotype			CG Genotype			GG Genotype			*P*-value
	N	Mean	SE	N	Mean	SE	N	Mean	SE	
Cord Blood	82	4.66	0.048	39	4.41	0.082	4	3.98	0.207	0.0010
Year 1	82	3.12	0.054	39	2.90	0.081	4	2.55	0.166	0.0089
Year 3	82	2.97	0.063	39	2.67	0.092	4	1.95	0.169	0.00025
Year 5	71	2.99	0.067	30	2.59	0.090	4	2.50	0.240	0.0016

Figure S1. Q-Q plot of the genome-wide *P*-values.

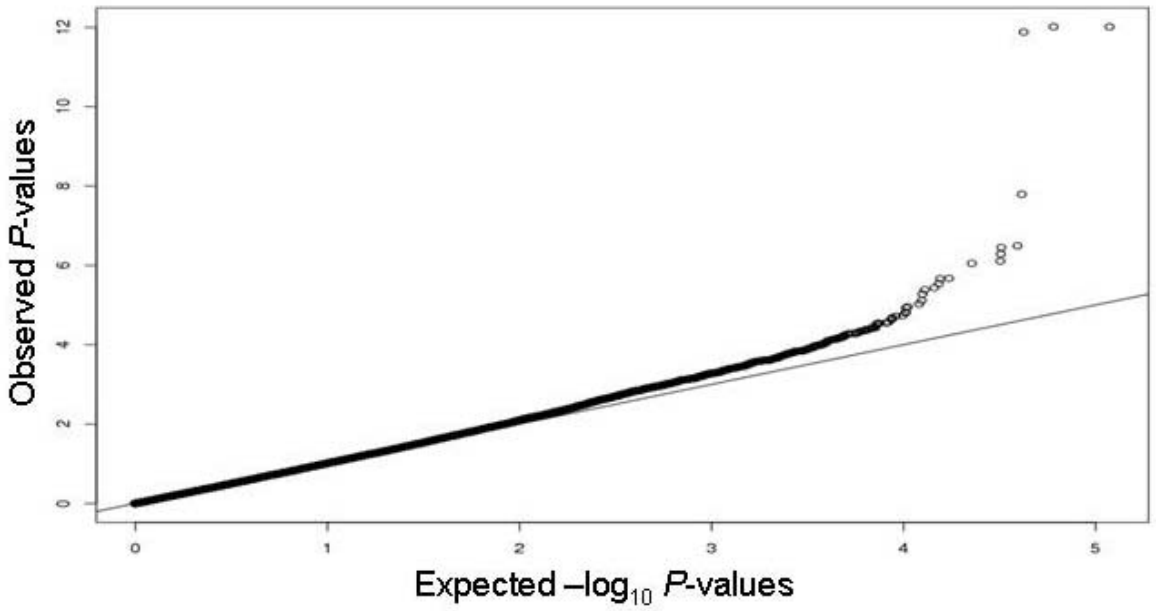

Figure S2. Genome-wide *P*-values for association with serum YKL-40 levels by chromosome.

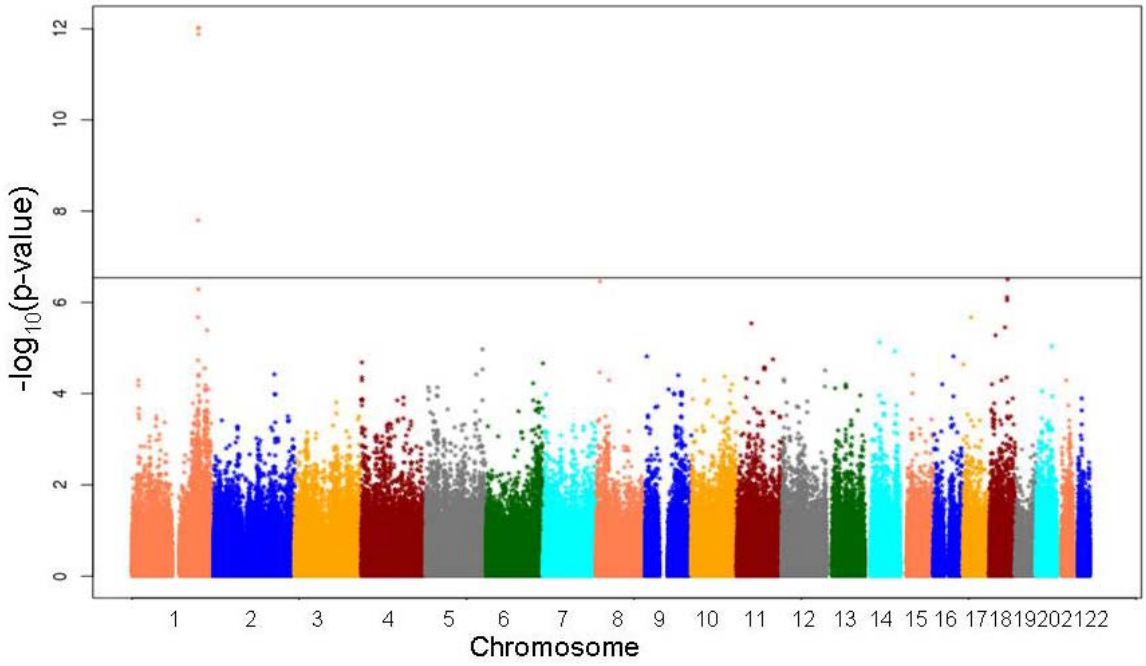

Figure S3. LD (r^2) between 10 SNPs genotyped in *CHI3L1* in the Hutterites (see Table 2 in paper) in 60 unrelated HapMap CEU (European) samples (upper left) and in 60 relatively unrelated (not first-degree relatives) Hutterites (lower right).