# A Second-Generation Association Study of the 5q31 Cytokine Gene Cluster and the Interleukin-4 Receptor in Asthma

Paula Kauppi,[1,2] Kerstin Lindblad-Toh,[6] Petteri Sevon,[3] Hannu T. T. Toivonen,[4,7]
John D. Rioux,[6] Anu Villapakkam,[6] Lauri A. Laitinen,[1] Thomas J. Hudson,[6,8]
Juha Kere,[2,5,*] and Tarja Laitinen[2]

*[1]Department of Medicine, Helsinki University Central Hospital, FI-00290 Helsinki, Finland,*
*[2]Department of Medical Genetics, [3]Department of Computer Science, [4]Rolf Nevanlinna Institute, and*
*[5]Finnish Genome Center, University of Helsinki, FI- 00014 Helsinki,Finland*
*[6]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 02142,USA*
*[7]Nokia Research Center, FI-00045 Helsinki, Finland,*
*[8]Montreal General Hospital Research Institute, McGill University, Montreal, Quebec H3T 1C5 Canada*

*\*To whom correspondence and reprint requests should be addressed. Fax: 358-9-191 25478. E-mail: juha.kere@helsinki.fi.*

**We have analyzed a dense set of single-nucleotide polymorphisms (SNPs) and microsatellites spanning the T-helper cytokine gene cluster (interleukins 3, 4, 5, 9, and 13, interferon regulatory factor-1, colony-stimulating factor-2, and T-cell transcription factor-7) on 5q31 and the gene encoding the interleukin-4 receptor (*IL4R*) on 16p12 among Finnish families with asthma. As shown by haplotype pattern mining analysis, the number of disease-associated haplotype patterns differed from that expected for the 129Q allele polymorphism in *IL13* for high serum total immunoglobulin (Ig) E levels, but not for asthma. The same SNP also yielded the best haplotype associations. For *IL4R*, asthma-associated haplotype patterns, most spanning the S411L polymorphism, showed suggestive association. However, these haplotypes consisted of the major alleles for the intracellular part of the receptor and were very common among both patients and controls. The minor alleles 503P and 576R have been reported to be associated with decreased serum IgE levels and changes in the biological activity of the protein, especially when inherited together. In the Finnish population, these two polymorphisms segregated in strong linkage disequilibrium. Our data support previous findings regarding *IL4R*, indicating that 503P and 576R may act as minor protecting alleles for IgE-mediated disorders.**

**Key Words: asthma, atopy, haplotype pattern mining, polymorphism, association**

## INTRODUCTION

The chromosome 5q31 cytokine gene cluster has been the most studied candidate gene region for asthma-related traits. In addition to the genes encoding the T helper 2-type cytokines (the interleukin genes *IL3*, *IL4*, *IL5*, *IL9*, and *IL13*), there are also other immunologically active genes such as those encoding interferon regulatory factor-1 (*IRF1*), colony-stimulating factor-2 (*CSF2*), and T-cell transcription factor-7 (*TCF7*). Several studies have reported suggestive linkage to atopy and high serum total immunoglobulin (Ig) E levels [1–3], bronchial hyperresponsiveness [4], and blood eosinophilia [5,6], and association to asthma [7], but so far the results have remained mutually inconclusive. An international effort was made to define the markers spanning the 5q cytokine cluster by linkage analysis in a total of 11 retrospectively pooled data sets (1037 families). The study showed suggestive linkage for asthma (LOD = 2.6) but no evidence for atopy [8].

Although it is not known whether human genes in the 5q31 cytokine cluster are genetic regulators for atopic disorders, experimental models of asthma have shown the importance of both IL4 and IL13 signaling. Administration of either exogenous IL4 or IL13 induced the asthma phenotype in mice [9]. However, neither of these cytokines was able to induce the asthma phenotype in mice deficient in the IL4 receptor (IL4R) [9]. The IL4 receptor is a heterodimer consisting of a common γ-chain that is shared by several other interleukin receptors (IL2, IL7, IL9, and IL13) and a ligand-specific α-chain encoded by *IL4R*. An IL4Rα–IL13Rα heterodimer is also able to transduce IL13 signaling [10] and is an important signaling pathway in the asthma model. All this indicates the importance of IL4Rα in IL4/IL13-mediated signaling.

**TABLE 1:** The genotype distribution of 5q31 cytokine cluster SNPs and *IL4R* SNPs among asthma patients and unaffected family based controls

| | Asthma genotypes | | | | | | Control genotypes | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *1 1* | | *1 2* | | *2 2* | | *1 1* | | *1 2* | | *2 2* | |
| | n | (%) | n | (%) | n | (%) | n | (%) | n | (%) | n | (%) |
| **5q31** | | | | | | | | | | | | |
| IL4ex1 | 14 | (9) | 92 | (57) | 55 | (34) | 9 | (7) | 75 | (58) | 46 | (35) |
| IL13ex4.2 | 113 | (91) | 11 | (9) | 0 | (0) | 93 | (93) | 7 | (7) | 0 | (0) |
| IL13ex4.1 (R129Q) | 64 | (39) | 82 | (50) | 17 | (11) | 62 | (47) | 51 | (39) | 19 | (14) |
| IL5pro | 12 | (7) | 67 | (41) | 84 | (52) | 10 | (8) | 59 | (44) | 64 | (48) |
| IRF1pro | 5 | (5) | 17 | (17) | 79 | (78) | 5 | (6) | 12 | (15) | 64 | (79) |
| CSF2enh1 | 0 | (0) | 5 | (3) | 159 | (97) | 0 | (0) | 2 | (1) | 132 | (99) |
| CSF2enh2 | 81 | (50) | 74 | (46) | 7 | (4) | 63 | (47) | 53 | (40) | 17 | (13) |
| CSF2ex4 | 33 | (23) | 64 | (46) | 43 | (31) | 25 | (20) | 54 | (44) | 43 | (35) |
| IL3 (P27S) | 11 | (9) | 68 | (54) | 47 | (37) | 17 | (15) | 49 | (42) | 50 | (43) |
| TCF7 | 15 | (10) | 61 | (39) | 80 | (51) | 5 | (4) | 57 | (44) | 66 | (52) |
| IL9int4 | 124 | (96) | 5 | (4) | 0 | (0) | 97 | (94) | 6 | (6) | 0 | (0) |
| IL9ex5 (T113M) | 2 | (1) | 31 | (22) | 109 | (77) | 4 | (3) | 25 | (21) | 89 | (76) |
| **IL4R** | | | | | | | | | | | | |
| I50V | 73 | (49) | 67 | (45) | 10 | (6) | 57 | (46) | 51 | (41) | 17 | (13) |
| Sil676 | 1 | (1) | 17 | (10) | 146 | (89) | 1 | (1) | 22 | (16) | 113 | (83) |
| Sil1114 | 0 | (0) | 1 | (1) | 163 | (99) | 0 | (0) | 0 | (0) | 136 | (100) |
| E375A | 118 | (85) | 21 | (15) | 0 | (0) | 100 | (85) | 17 | (15) | 0 | (0) |
| Sil1417 | 128 | (81) | 28 | (18) | 1 | (1) | 102 | (77) | 29 | (22) | 2 | (1) |
| C406R | 118 | (85) | 20 | (14) | 1 | (1) | 95 | (78) | 25 | (20) | 2 | (2) |
| S411L | 113 | (98) | 2 | (2) | 0 | (0) | 87 | (94) | 5 | (5) | 1 | (1) |
| S503P | 124 | (83) | 21 | (14) | 5 | (3) | 83 | (72) | 24 | (21) | 8 | (7) |
| Q576R | 121 | (72) | 41 | (25) | 5 | (3) | 82 | (63) | 37 | (28) | 11 | (8) |

*IL4R*, located on chromosome 16p12, is exceptionally polymorphic. Eight missense and six silent polymorphisms have been described [11–13]. Some of the polymorphisms have been reported to be associated with IgE-mediated disorders [12–14]. Although the evidence for association is preliminary and contradictory results exist [15,16], those findings are of interest, as the reported single-nucleotide polymorphisms (SNPs) may change the function of the molecule. The amino acid change from valine to isoleucine in position 50, which was associated with atopic asthma in the Japanese population, upregulated the receptor response to IL4 when transfected into both mouse and human cell lines [14]. The 576R allele, which has been reported to be associated with atopy and hyper-IgE syndrome, was also associated with higher levels of expression of the low-affinity receptor for IgE (CD23) than the wild type allele after IL4 induction [12]. However, 576R, when present together with

another polymorphism that alters an amino acid in the intracellular part of *IL4R* (P503), is associated with lowered serum IgE production, possibly because of conformational changes in the receptor that influence signal transduction pathways [16].

If susceptibility alleles are common in the general population and have rather small effects, association can be a more sensitive method than linkage to identify these loci [17]. Our previous study of Finnish asthma patients using microsatellites from 5q31–q33 failed to show evidence of a haplotype association [18]. The increasing knowledge of the human genome and the large number of SNPs that are becoming available as well as improved technology for genotyping make largescale association studies possible. Also, association studies such as that describing a SNP in the promotor (–159C→T) of *CD14* and an association with IgE level [19], and another showing the amino acid substitution R129Q in *IL13* and association with asthma [20] have given us new, useful data compared with microsatellite-based association studies, with which there is no possibility of functional testing. In addition, simultaneous genotyping of many genes also allows study of specific two-gene allele combinations in IL4–IL4R and IL13–IL4R signaling.

# RESULTS

## Linkage Analysis

We analyzed both asthma and high serum total IgE levels as qualitative traits using multipoint nonparametric linkage analysis among the Finnish asthma families. We used 42 microsatellites together with 12 SNPs for chromosome 5, and 11 microsatellites with 9 SNPs for chromosome 16 placed evenly across the chromosome [18,21]. Neither of the candidate regions showed any evidence of linkage, confirming our previous observations [18]. For the 5q cytokine cluster, the nonparametric linkage score was negative, varying from –0.3 to –1.9 for both phenotypes. For the *IL4R* SNPs, the nonparametric linkage score varied from –1.4 to 0.1 for both phenotypes.
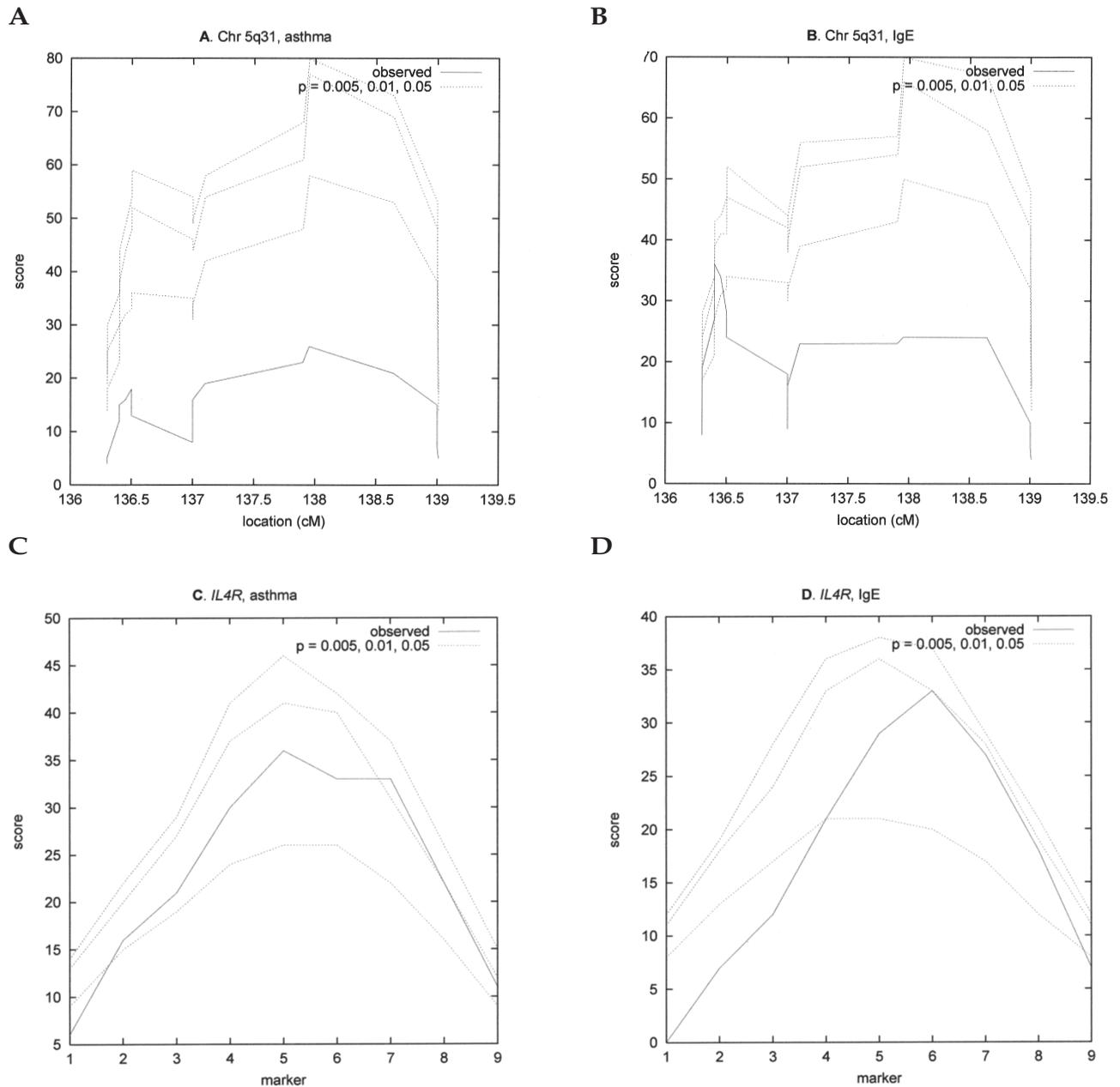
**FIG. 1.** Results of observed (continuous line) and permutated (dotted lines) HPM data on the 5q31 cytokine cluster and *IL4R* in asthma and high IgE levels. (A–D) The permuted data sets ($n = 1000$) were organized according to the best score. Dotted lines show the fifth-best score ($P = 0.005$), the tenth-best score ($P = 0.01$), and the fiftieth-best score ($P = 0.05$).

## Allele Associations

We compared the genotype frequencies of the SNPs in the 5q31 cytokine cluster and in *IL4R* among asthma patients (one affected person randomly selected from each family) with those among the family-based controls (Table 1). The distribution of alleles did not differ between the study groups. We studied the serum total IgE level as a dichotomized trait among individuals with either high (IgE >100 kU/L) or low (IgE {≤}100 kU/L) IgE levels. We found

no deviation (data not shown). We analyzed genotype frequencies in *IL4R* in a similar way (Table 1) and again found no allele associations. Alleles for all markers were in Hardy–Weinberg equilibrium and the overall success rate for SNP genotyping was 83%.

## Haplotype Association Analysis

We selected 12 SNPs together with five previously genotyped microsatellites in the 5q31 cytokine cluster spanning 3.1 Mb

**TABLE 2:** Selected haplotypes of all trios of chromosome 5q cytokine cluster region obtained by data mining for asthma phenotype

| Total | All | Asthma | Conf(%) | $\chi^2$ | IL4 | IL4ex1 | IL13ex4.2 | IL13ex4.1 (R129Q) | IL5pro | IRF1pro | IRF1 | CSF2ex4 | CSF2enh2 | CSF2enh1 | IL3 (P27S) | TCF7 | D5S1995 | D5S2115 | IL9int4 | IL9 | IL9ex5 (T113M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 357 | 21 | 17 | 81 | 5.3 | - | - | - | - | 1 | 2 | 5 | 2 | 1 | - | - | - | - | - | - | - | - |
| 390 | 69 | 49 | 71 | 6.8 | - | - | - | - | - | 2 | 5 | 2 | 1 | - | - | - | - | - | - | - | - |
| 380 | 65 | 46 | 71 | 6.2 | - | - | - | - | - | 2 | 5 | 2 | 1 | 2 | - | - | - | - | - | - | - |
| 421 | 41 | 31 | 76 | 7.7 | - | - | - | - | - | - | - | 1 | 2 | 1 | 2 | 7 | - | - | - | - | - |

| Total | All | High IgE | Conf(%) | $\chi^2$ | IL4* | IL4ex1 | IL13ex4.2 | IL13ex4.1 (R129Q) | IL5pro | IRF1pro | IRF1* | CSF2ex4 | CSF2enh2 | CSF2enh1 | IL3 (P27S) | TCF7 | D5S1995* | D5S2115* | IL9int4 | IL9* | IL9ex5 (T113M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 352 | 123 | 79 | 65 | 5.2 | - | - | - | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 274 | 77 | 57 | 74 | 10.8 | - | - | 1 | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 217 | 30 | 24 | 80 | 6.9 | - | 1 | 1 | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 301 | 18 | 16 | 94 | 10.0 | 5 | 2 | 1 | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 301 | 15 | 14 | 93 | 7.8 | 15 | 2 | 1 | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 301 | 15 | 14 | 93 | 7.8 | 17 | 2 | 1 | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 209 | 64 | 47 | 73 | 5.8 | - | - | - | 2 | 2 | 2 | - | - | - | - | - | - | - | - | - | - | - |
| 259 | 11 | 9 | 91 | 5.1 | - | - | - | 2 | 2 | 2 | 1 | 2 | - | - | - | - | - | - | - | - | - |

*microsatellite

Total, number of studied chromosomes; All, number of haplotypes with a specific pattern; Asthma/ High IgE, number of disease associated haplotypes with a specific pattern; Conf, percentage of haplotypes with the specific pattern that are associated with disease; $\chi^2$, chi square value for disease association of the specific haplotype pattern. For the haplotype patterns 30% of missing data but no errors were allowed.

and 9 SNPs in *IL4R* spanning 24 kb [13] for the haplotype analyses in 186 trios. On chromosome 5, the trios showed 299 affected and 245 control chromosomes for asthma and 231 affected and 177 control chromosomes for high IgE levels. For chromosome 16, the trios consisted of 320 affected and 260 control chromosomes for asthma and 257 affected and 195 control chromosomes for high IgE levels.

Tables 2 and 3 show the best associating haplotype clusters in 5q and *IL4R* for asthma and high IgE levels when 30% missing data but no errors were allowed in the haplotype patterns. Most of the asthma 5q haplotypes spanned the marker CSFenh2. Extension of the haplotype in either direction from CSFenh2 failed to produce better confidence values (Table 2), indicating that CSFenh2 did not contain disease-associated haplotypes. For high IgE levels, extension of the haplotype around IL13ex4.1 (129Q) did lead to better disease association, but the $\chi^2$ values were rather low and did not allow much cor-

rection for multiple testing. For *IL4R* (Table 3), haplotypes associated with asthma and those associated with high IgE levels showed a cluster of haplotypes around Sil1114 and S411L. For Sil1114, all associations were weak. 411S (the major allele) reached the best association for both asthma ($\chi^2 = 7.0$) and high IgE levels ($\chi^2$ value = 3.8), but again, none of the extended haplotypes reached better association or confidence to either of the phenotypes. Moreover, the neighboring markers in the extended haplotypes contained the major alleles in the population. Analysis of the minor alleles of the coding polymorphisms located in the intracellular part of the receptor showed that 503P and 576R segregated in strong linkage disequilibrium ($P = 3.2 \times 10^{-46}$). They formed part of a haplotype (406C–411L–503P–576R) that showed reverse association with asthma (12 control versus 0 affected chromosomes; $\chi^2$, 14.7) and high IgE levels (10 control versus 2 affected chromosomes; $\chi^2$, 6.1).

**TABLE 3:** Selected haplotypes of all trios of *IL4R* obtained by data mining for high IgE phenotype

| Total | All | Asthma | Conf (%) | $\chi2$ | I50V | Sil676 | Sil1114 | E375A | Sil1417 | C406R | S411L | S503P | Q576R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 364 | 323 | 193 | 60 | 1.8 | - | - | - | - | - | 1 | 1 | - | - |
| 433 | 422 | 245 | 58 | 7.0 | - | - | - | - | - | - | 1 | - | - |
| 388 | 337 | 201 | 60 | 2.1 | - | - | - | - | - | - | 1 | 1 | - |
| 370 | 306 | 183 | 60 | 2.2 | - | - | - | - | - | - | 1 | 1 | 1 |
| 360 | 311 | 187 | 60 | 2.3 | - | - | - | - | - | 1 | 1 | 1 | - |
| 458 | 374 | 221 | 59 | 5.4 | - | - | - | - | - | 1 | 1 | 1 | 1 |
| 462 | 393 | 229 | 58 | 5.2 | - | - | - | - | 1 | 1 | 1 | 1 | - |
| 452 | 371 | 218 | 59 | 4.8 | - | - | - | - | 1 | 1 | 1 | 1 | 1 |
| 414 | 284 | 165 | 58 | 3.6 | 1 | 2 | - | - | - | - | - | - | - |
| 409 | 281 | 163 | 58 | 2.3 | 1 | 2 | 2 | - | - | - | 1 | 1 | 1 |
| 544 | 501 | 283 | 57 | 4.7 | - | 2 | 2 | - | - | - | - | - | - |
| 418 | 30 | 20 | 70 | 2.3 | - | 2 | 2 | 2 | - | - | - | - | - |
| 428 | 30 | 20 | 70 | 2.4 | - | - | 2 | 2 | - | - | - | - | - |
| 409 | 25 | 19 | 76 | 4.0 | - | - | 2 | 2 | 2 | - | 1 | - | - |
| Total | All | High IgE | Conf(%) | $\chi2$ | I50V | Sil676 | Sil1114 | E375A | Sil1417 | C406R | S411L | S503P | Q576R |
| 264 | 236 | 144 | 61 | 0.6 | - | - | - | - | - | 1 | 1 | - | - |
| 317 | 309 | 182 | 59 | 3.8 | - | - | - | - | - | - | 1 | - | - |
| 270 | 238 | 144 | 61 | 0.7 | - | - | - | - | - | - | 1 | 1 | - |
| 262 | 13 | 10 | 77 | 1.6 | - | - | - | - | - | - | 1 | 1 | 2 |
| 308 | 75 | 44 | 60 | 0.4 | 2 | 2 | - | - | - | - | - | - | - |
| 423 | 393 | 226 | 58 | 2.3 | - | 2 | - | - | - | - | - | - | - |
| 422 | 391 | 223 | 57 | 1.7 | - | 2 | 2 | - | - | - | - | - | - |
| 322 | 276 | 163 | 59 | 0.4 | - | 2 | 2 | 1 | - | - | - | - | - |

Total, number of studied chromosomes; All, number of haplotypes with a specific pattern; Asthma/High IgE, number of disease-associated haplotypes with a specific pattern; Conf, percentage of haplotypes with the specific pattern that are associated with disease; $\chi2$, chi square value for disease association of the specific haplotype pattern. For the haplotype patterns 30% of missing data but no errors were allowed.

## Haplotype Pattern Mining

We used the following parameters for haplotype pattern mining (HPM) analysis: maximum pattern length, five markers; allowed number of gaps, one (for missing data and possible errors); and $\chi2$ threshold for the association, 3.0. Continuous lines in Fig. 1 show the scores (number of qualified haplotype patterns spanning across the marker) for each marker and phenotype. For asthma, we found no significant associations in 5q31 (best $P$ value, 0.28 at the marker IL13ex4.1; Fig. 1A). The best $P$ value (0.015) was obtained at the marker IL13ex4.1 (R129Q) for high IgE level (Fig. 1B). In *IL4R*, for high IgE level, the best $P$ value (0.010) was at the marker C406R (Fig. 1D), and for asthma, at the marker S411L ($P = 0.007$; Fig. 1C). To analyze the significance of our best observation, we used a permutation test. We randomized the grouping of chromosomes 1000 times. We then analyzed these permutated data sets using identical HPM settings. Dotted lines in Fig. 1 show the scores of the permutated data at $P$ levels 0.005, 0.01, and 0.05. The best $P$ values were between 0.005 and 0.01. To evaluate the overall significance of the smallest marker-wise $P$ value, it must be corrected for multiple testing that is dependent on the number of loci tested ($n = 2$), the number of markers tested ($n = 24$), the number of phenotypes tested ($n = 2$), and the number of HPM settings tested ($n = 9$). As most of the variables are highly correlated, a simple Bonferroni correction is probably too conservative. Correction for two chromosomal regions and for two partially correlated phenotypes give, by simulation, a minimum factor of 2.6 for the correction (data not shown). The effective number of independent tests estimating the effect of nearby markers or different HPM settings is more difficult to determine. However, for most of the $P$ values below 0.05, the significance vanished with this correction. Even the smallest $P$ value (0.007) left little "headroom" for further correction. However, based on these results, it is possible that the observed associations near IL13ex4.1 for high IgE levels or in IL4R for asthma and for high IgE levels remain significant at 95% probability level.

To study the overall divergence of different *IL4R* haplotypes in the study population, we analyzed 310 unrelated chromosomes that could be determined unambiguously for all nine markers genotyped. In these chromosomes we found 22 different haplotypes. However, three haplotypes covered 84% of all chromosomes. *IL4R*\*1 2 2 1 1 1 1 1 1 was the most common haplotype, accounting for 57% of the haplotypes. *IL4R*\*2 2 2 1 1 1 1 1 1 accounted for 23%, and *IL4R*\*2 1 2 1 1 1 1 1 1, for 4% of the chromosomes.

### Allele Combinations in IL4–IL4R and IL13–IL4R Signaling Pathways

As the 5q31 cytokines and *IL4R* are located on different chromosomes, both loci are inherited independently. Any deviation from random allele distribution among the patients would indicate selection that favors certain combinations. We analyzed the allele combinations of *IL4–IL4R* and of *IL13–IL4R* in each asthma patient by a $2 \times 2$ contingency table. To avoid ascertainment bias, we randomly selected only one affected family member from each family. Because our sample size did not have sufficient statistical power to evaluate the significance of very rare combinations, we excluded affected individuals who were homozygous for the minor alleles in both marker loci. None of the allelic combinations differed significantly from the expected result (data not shown).

## DISCUSSION

Here we have demonstrated allele and haplotype associations among Finnish asthma families with a dense set of SNPs and microsatellite markers in the cytokine gene cluster on 5q31 and in *IL4R* on 16p12. On 5q31, the 129Q variant of *IL13* was the only marker that showed any evidence for association with either of the studied phenotypes. An excess of this allele has been reported among the Japanese and British atopic asthma patients [20]. Use of a partly identical set of *IL4R* markers demonstrated significant allele and haplotype associations in three ethnic groups [13]. However, because the results differed between the study groups, the susceptibility allele(s) could not be named [13]. Our HPM analysis results were of borderline statistical significance, but the four last missense polymorphisms in the intracellular part of the receptor seem to be important. The susceptibility haplotype contained the major alleles of all four markers, making it difficult to interpret the biological relevance of the observed haplotype association. The minor alleles, 411L–503P–576R, showed some evidence of protective effect for asthma, supporting observations made previously [16]. Because weak genetic regulators can be of great importance in certain combinations, we have also reported two-locus association studies of ligand–receptor gene pairs. None of the allele combinations in IL4–IL4R or in IL13–IL4R signaling was favored among the study groups.

Compared with our previous study of the 5q31–q33 region [18], the marker density in the cytokine gene cluster was higher here and the physical positions of the markers were better defined. Microsatellites were added to increase the haplotype information. For *IL4R*, we found only three major haplotypes that covered 84% of the chromosomes. Comparing this rather uniform distribution with previously reported distributions using in part the same markers (I50V–E375A–C406R–S503P–Q576R), both the Kainuu and the Hutterite populations showed less genetic divergence than the mixed Caucasians or African Americans [13].

For haplotype analysis we used the HPM method, which is powerful in locating a disease-causing gene when one is known to exist [22]. By allowing gaps in haplotype patterns, it is also more robust for genotyping errors, marker mutations, unrecognized recombinations, and missing data than the haplotype analysis we used in our first study [18,22]. Also, the HPM parameters can be optimized for each data set by studying individual features of genotype data such as marker information, marker density, and missing genotypes. In our study, few haplotype associations could be detected in our initial analysis (Tables 2 and 3), but the $\chi^2$ values remained rather low. In conclusion, our re-analysis of the 5q31 cytokine gene cluster has shown that, among the Finnish asthma families, 129Q in IL13 and the intracellular SNPs of *IL4R* are potential minor genetic regulators in the development of IgE-mediated disorders.

## MATERIALS AND METHODS

*Study population.* The study population consisted of 162 families (324 asthma patients and their 358 non-affected family members) recruited based on asthma from central eastern Finland (Kainuu province). Probands were recruited as having self-reported, physician-diagnosed asthma. The diagnoses and the diagnostic testing were verified from the medical records as described [23]. A physician interviewed all patients and they filled out a questionnaire about their respiratory symptoms and allergies. We measured serum total IgE levels in all participants using Diagnostics CAP FEIA (Kabi Pharmacia, Uppsala, Sweden ). Of the families studied, 79 were multiplex pedigrees consisting of 69 affected sibling pairs, 28 affected cousin pairs, and 42 other affected relative pairs. The families were included for linkage analysis without further selection. The families studied included 83 nuclear trios (father, mother, and child) with one or two affected family members (never both parents). These families were included in the association study together with trios in the multiplex families ($n = 103$). From multiplex families, only those non-overlapping trios with full genotype information for father, mother, and offspring were accepted.

We studied two phenotypes: asthma and IgE responsiveness dichotomized as high (> 100 kU/l) or low ($\leq$ 100 kU/l) serum total IgE level [4,24]. Both the probands whose diagnosis could not be verified and the family members who reported asthma-like symptoms were classified as having an unknown phenotype. The study was approved by the ethical committees of the Kainuu Central Hospital and Department of Medical Genetics, University of Helsinki, Finland.

*SNP genotyping by single-base-pair extension.* We extracted DNA from peripheral blood leucocytes using a standard non-enzymatic method. We genotyped 12 SNPs from the chromosome 5q31 cytokine gene cluster: IL4ex1 (32,711T174C in clone AC004039), IL13ex4.1 (46,457G174A in clone AC004039), IL13ex4.2 (45,977G174A in clone AC004039), IL5pro (37,651T174C in clone AC004039), IRF1pro (1920T174C in clone AC002194), CSF2enh1 (14,259G174T in clone AC004511), CSF2enh2 (14,348T174G in clone AC004511), CSF2ex4 (19,375C174T in clone AC004511), IL3 (4393T174C in clone AC004511), TCF7 (1581T174C in clone L14482), IL9int4 (125,992A174G in clone AC002428), and IL9ex5 (127,462T174C in clone AC002428) (details available at

**Article**

http://www.genome.helsinki.fi/english/research/projects.xml; John D. Rioux, unpublished data). Of the SNPs studied, six were exonic and one was intronic; three caused an amino acid change in IL13, IL3, and IL9; two were upstream of IL5 and IRF1; and two were upstream of CSF2. Markers were located using available sequence (http://www.ncbi.nlm.nih.gov), the physical map of human chromosome 5 (http://www-hgc.lbl.gov/human-maps.html), and Marshfield's genetic map (http://www.marshmed.org/genetics). For IL4R on chromosome 16p12, seven missense and three silent SNPs were screened: I50V, E375A, C406R, S411L, S503P, Q576R, S761P, SiI676T174C, SiI1114C174T, and SiI1417G174T (details available at http://www.genome.helsinki.fi/english/research/projects.xml) [11,16].

For SNP genotyping, we used length-multiplexed single-base extension as described [25]. PCR primers were designed to be as similar as possible to the SNP (maximum length, 150 bp). Loci were subjected to two rounds of PCR amplification. In the primary amplification, the forward primers had T7 tails at the 5′ ends and the reverse primers had T3 tails at the 5′ ends. Primer pairs were checked for homology to all amplicons and sorted into pools consisting of up to 50 primer pairs. In the first round, 10 ng genomic DNA was amplified with a pool of primer pairs (0.1 μM) and 2.5 units of Amplitaq Gold (Perkin Elmer, Norwalk, CT). In the second round, a 3-μl aliquot of the primary amplification product was amplified with biotinylated T7 and biotinylated T3 primers. A 7-μl aliquot of this secondary amplification product was purified from the unincorporated dNTPs using streptavidin-coated Dynabeads (Dynal, Oslo, Norway). We then did multiplex SBE reaction on the purified product using SNP-specific primers, 2′,7′-dimethoxy-4′,5′-dichloro-6-carboxyfluorescein (JOE)-labeled ddATP (0.12 M), $N,N,N',N'$-tetramethyl-6-carboxyrhodamine (TAMRA)-labeled ddCTP (0.12 M), 5-carboxyfluorescein (FAM)-labeled ddGTP (0.12 M), 6-carboxy-X-rhodamine (ROX)-labeled ddUTP (0.60 M; NEN DuPont, Boston, MA), and 0.5 Thermosequenase (Amersham, Arlington Heights, IL). Excess ddNTPs were removed from the SBE products using 96-well gel filtration blocks (Edge Biosystems) before samples were separated by electrophoresis on ABI 377 sequencers. We analyzed the length-multiplexed single-base extension gels using a system developed at the Whitehead Institute/MIT Center for Genome Research [25].

*SNP genotyping by specific restriction enzyme digestion sites.* To validate our genotyping method, we also genotyped I50V, E375A and C406R, S503P, Q576R, and S761P using restriction enzyme digestion. For the primer sequences and restriction enzymes, detailed information is available (http://www.genome.helsinki.fi/english/research/projects.xml). The reaction products were cleaved using restriction enzymes (BioLabs, Boston, MA), separated by electrophoresis through agarose gels and photographed under ultraviolet illumination. S761P showed no polymorphism in our sample set and was excluded from the analysis.

*Haplotyping.* Large pedigrees were divided into trios using an in-house computer program. The program identifies the maximal number of trios that are not overlapping and in which one or two members were affected (not both parents). Trios that included members who had not been genotyped or members with unknown phenotype were excluded. Haplotyping was done within each trio, and from each trio, four independent chromosomes were obtained. For ambiguities (missing genotype data, identical heterozygotic genotypes in all of the family members, or Mendel errors), the alleles were discarded. If the child was affected, the transmitted chromosomes were considered disease-associated and the nontransmitted chromosomes, as controls. If one of the parents was affected, his or her chromosomes were considered disease-associated and the spouse's chromosomes, as controls. If both the parent and the child were affected, only the nontransmitted chromosome of an unaffected parent was considered as the control and the other three, as disease-associated. These haplotypes were used as input for HPM [22].

*Statistical analysis.* Both chromosomes 5 and 16 were analyzed for linkage using the computer package GENEHUNTER 2.0 [26]. The SNP genotypes were combined with 42 polymorphic microsatellites for chromosome 5 and with 11 microsatellites for chromosome 16 from our genome scan [21].

All allele and haplotype comparisons were made using the $\chi^2$ test except when the expected number in a single cell in a 2 × 2 contingency table was less than five; in these cases, Fisher's exact test was applied. To correct for multiple testing, nominal $P$ values were multiplied by the number of tests.

Haplotype analysis was done using HPM [22]. The input data consisted of chromosomes tagged with the disease status (affected or unaffected). To make the algorithm more robust for marker mutations, genotyping errors, and missing data, gaps were allowed in the patterns. The maximum length of the haplotype patterns and the number of gaps allowed were given as parameters for the algorithm. A haplotype pattern matches the chromosome when all of its non-gap alleles agree with the respective alleles in the chromosome.

The algorithm finds all the haplotype patterns that are strongly disease-associated by a 2 × 2 $\chi^2$ test. The threshold that the $\chi^2$ value must exceed can be given as a parameter. Each marker is assigned a score with the number of qualified patterns spanning across the marker. The method relies on the assumption that in the proximity of a disease susceptibility gene there is more linkage disequilibrium in the set of disease-associated chromosomes than in the set of nonassociated chromosomes. The greater the difference in levels of linkage disequilibrium, the more significant the disease-associated patterns found.

The scores of different markers are not directly comparable, as marker densities and their information content vary. To compensate for this, HPM uses randomization to obtain marker-wise $P$ values that are comparable with each other. At each randomization cycle, the disease association status of each haplotype is assigned at random, keeping the total number of affected and unaffected haplotypes constant and the scores are re-calculated. The $P$ value for a given marker is the proportion of iterations at which a score larger or equivalent to the experimental pattern is obtained. The disease susceptibility gene is most likely to lie in the proximity of the marker with the lowest $P$ value [22].

## REFERENCES

1. Marsh, D. G., *et al.* (1994). Linkage analysis of IL4 and other chromosome 5q31.1 markers and total serum immunoglobulin E concentrations. *Science* **264:** 1152–1156.

2. Meyers, D. A., *et al.* (1994). Evidence for a locus regulating total serum IgE levels mapping to chromosome 5. *Genomics* **23:** 464–470.

3. Xu, J., *et al.* (2000). Major genes regulating total serum immunoglobulin E levels in families with asthma. *Am. J. Hum. Genet.* **67:** 1163–1173.

4. Postma, D. S., *et al.* (1995). Genetic susceptibility to asthma—bronchial hyperresponsiveness coinherited with a major gene for atopy. *N. Engl. J. Med.* **333:** 894–900.

5. Martinez, F. D., *et al.* (1998). Linkage of circulating eosinophils to markers on chromosome 5q. *Am. J. Respir. Crit. Care Med.* **158:** 1739–1744.

6. Rioux, J. D., *et al.* (1998). Familial eosinophilia maps to the cytokine gene cluster on human chromosomal region 5q31-q33. *Am. J. Hum. Genet.* **63:** 1086–1094.

7. Noguchi, E., *et al.* (1998). Association of asthma and the interleukin-4 promoter gene in Japanese. *Clin. Exp. Allergy* **28:** 449–453.

8. Lonjou, C., *et al.* (2000). A first trial of retrospective collaboration for positional cloning in complex inheritance: Assay of the cytokine region on chromosome 5 by the Consortium on Asthma Genetics (COAG). *Proc. Natl. Acad. Sci. USA* **97:** 10942–10947.

9. Grunig, G., *et al.* (1998). Requirement for IL-13 independently of IL-4 in experimental asthma. *Science* **282:** 2261–2263.

10. Izuhara, K., and Shirakawa, T. (1999). Signal transduction via the interleukin-4 receptor and its correlation with atopy. *Int. J. Mol. Med.* **3:** 3–10.

11. Deichmann, K., Bardutzky, J., Forster, J., Heinzmann, A., and Kuehr, J. (1997). Common polymorphisms in the coding part of the IL4-receptor gene. *Biochem. Biophys. Res. Commun.* **231:** 696–697.

12. Hershey, G. K., Friedrich, M. F., Esswein, L. A., Thomas, M. L., and Chatila, T. A. (1997). The association of atopy with a gain-of-function mutation in the α subunit of the interleukin-4 receptor. *N. Engl. J. Med.* **337:** 1720–1725.

13. Ober, C., *et al.* (2000). Variation in the interleukin 4-receptor α gene confers susceptibility to asthma and atopy in ethnically diverse populations. *Am. J. Hum. Genet.* **66:** 517–526.

14. Mitsuyasu, H., *et al.* (1998). Ile50Val variant of IL4R α upregulates IgE synthesis and associates with atopic asthma. *Nat. Genet.* **19:** 119–120.

15. Noguchi, E., *et al.* (1999). Lack of association of atopy/asthma and the interleukin-4 receptor α gene in Japanese. *Clin. Exp. Allergy* **29:** 228–233.

16. Kruse, S., *et al.* (1999). The polymorphisms S503P and Q576R in the interleukin-4 receptor α gene are associated with atopy and influence the signal transduction. *Immunology* **96:** 365–371.

17. Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* **273:** 1516–1517.

18. Laitinen, T., *et al.* (1997). Genetic control of serum IgE levels and asthma: linkage and

linkage disequilibrium studies in an isolated population. *Hum. Mol. Genet.* **6:** 2069–2076.

**19.** Baldini, M., *et al.* (1999). A polymorphism in the 5′ flanking region of the CD14 gene is associated with circulating soluble CD14 levels and with total serum immunoglobulin E. *Am. J. Respir. Cell Mol. Biol.* **20:** 976–983.

**20.** Heinzmann, A., *et al.* (2000). Genetic variants of IL-13 signalling and human asthma and atopy. *Hum. Mol. Genet.* **9:** 549–559.

**21.** Laitinen, T., *et al.* (2001). A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat. Genet.* **28:** 87–91.

**22.** Toivonen, H. T. T., *et al.* (2000). Data mining applied to linkage disequilibrium mapping. *Am. J. Hum. Genet.* **67:** 133–145.

**23.** Kauppi, P., Laitinen, L. A., Laitinen, H., Kere, J., and Laitinen, T. (1998). Verification of self-reported asthma and allergy in subjects and their family members volunteering for gene mapping studies. *Resp. Med.* **92:** 1281–1288.

**24.** Burrows, B., Martinez, F. D., Halonen, M., Barbee, R. A., and Cline, M. G. (1989). Association of asthma with serum IgE levels and skin-test reactivity to allergens. *N. Engl. J. Med.* **320:** 271–277.

**25.** Lindblad-Toh, K., *et al.* (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.* **24:** 381–386.

**26.** Kruglyak, L., Daly, M. J., Reeve-Daly, M. P., and Lander, E. S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58:** 1347–63.