J Hum Genet (2002) 47:684–687

SHORT COMMUNICATION

Rana Nagarkatti · Balaram Ghosh

Identification of single-nucleotide and repeat polymorphisms in two candidate genes, interleukin 4 receptor (*IL4RA*) and signal transducer and activator of transcription protein 6 (*STAT6*), for Th2-mediated diseases

Received: August 5, 2002 / Accepted: September 30, 2002

Abstract We report here the identification of 26 single-nucleotide polymorphisms (SNPs) spanning a total of 147 kb in two candidate genes, *IL4RA* and *STAT6*, for atopic disorders. Fourteen novel SNPs were found in our population. We also report the identification of three novel polymorphic (CA) repeat regions in these genes. No insertion/deletion polymorphisms in *AluY* elements were detected. The encoded proteins of these two genes are part of a single signaling pathway, and therefore, functional polymorphisms in these genes could potentially lead to higher risk and susceptibility to atopic disorders. We also examined the allelic frequency and haplotypes of these polymorphisms in a control population. These data will be potentially useful for association studies designed to investigate the role of these genes in atopic disorders such as asthma, eczema, and allergic rhinitis. This is the first report on the polymorphic content of these two genes in the Indian population.

Key words Th2 · Atopy · *IL4RA* · *STAT6* · *AluY* · (CA) repeat · 5′ UTR · Indian population

Introduction

Atopy is the basis for the development of various diseases such as dermatitis, rhinitis, and asthma (Cookson 1999; Barnes and Marsh 1998). The T-helper type 2 (Th2) cells, which release mainly interleukin-4 (IL-4) and interleukin-5 (IL-5), are responsible for the deviation of the immune system toward atopicity, with increased total serum IgE levels (Biedermann and Rocken 1999). When IL-4 engages its receptor complex (IL4RA), the phosphorylation, dimerization, and nuclear localization of the signal transducer and activator of transcription 6 (STAT6) protein results in the transcription of the ε gene (Nelms et al. 1998; Izuhara et al. 2000). Accordingly, *STAT6* (−/−) mice failed to develop a Th2 response and demonstrated a lack of IgE production (Shimoda et al. 1996). Moreover, mice treated with IL4RA antagonists also showed a decrease in atopic phenotypes (Tomkinson et al. 2001).

Recently, various genetic studies have shown that polymorphisms in *IL4Rα* are associated with atopic disorders (Ober et al. 2000; Takabayashi et al. 2000). Similarly, the genomic region harboring the *STAT6* gene (12q21.2) has also been found to be associated with such disorders (Tamura et al. 2001). In addition, several functional polymorphisms that lead to aberrant expression and signal transduction resulted in atopic phenotypes (Hackstein et al. 2001; Mitsuyasu et al. 1999). In contrast, these genes were not found to be associated with atopy in a Caucasian and Japanese study (Duetsch et al. 2002; Tanaka et al. 2001). Thus, the positional candidate gene approach for dissection of such complex disorders will require the identification of informative single-nucleotide polymorphisms (SNPs) in various populations.

Here, we report the presence of polymorphisms (SNPs, simple repeats, and *AluY*) spanning a total of 147 kb in *IL4RA* and *STAT6* genes of the Indian population and the predicted haplotypes.

Subjects and methods

Unrelated atopic and control individuals were selected using a questionnaire for obtaining details pertaining to migration status, ethnicity, and clinical phenotyping

R. Nagarkatti · B. Ghosh (✉)
Molecular Immunogenetics Laboratory, Centre for Biochemical Technology, Mall Road, New Delhi 110007, India
Tel. +91-7666156, +91-7666157, ext 140; Fax +91-7667471, +91-7416489
e-mail: bghosh@cbt.res.in

Accession numbers for SNPs reported in this article have been deposited in dbSNP (http://www.ncbi.nih.nlm.gov/dbSNP) with the following IDs: ss#5105839, 5105840, 5105841, 5105842, 5105843, 5105844, 5105845, 5105846, 5105847, 5105848, 5105849, 5105850, 5105851, 5105852, 5105853, 5105854, 5105855, 5105856, 5105857, 5105858, 5105859, 5105860, 5105861, 5105862.

(Nagarkatti et al. 2002). Based on these parameters, all individuals were grouped as North Indian samples. The average age of the controls was 27.5 years (8.5 years) and the sex ratio was 0.59:0.41 (M:F). All individuals gave their informed consent and the institutional ethics committee approved the study. DNA was isolated using a standard protocol. Sequences were downloaded from National Centre for Biotechnology Information Entrez (*STAT6*, Accession number AH006951; *IL4RA*, Accession number AC004525). Primers were designed to screen by sequencing the complete coding regions, extensive parts of the intronic region, and the promoter region for putative polymorphic elements. A total length of 38.83 kbp was sequenced, encompassing a genomic region of 147 kbp (location and nucleotides of sequencing primers for IL4RA and STAT6 can be obtained from the author on request at bghosh@cbt.res.in). Putative repeats and *Alu Y* elements were identified using the RepeatMasker software (http://ftp.genome.washington.edu/RM/RepeatMasker.html). Repeats were genotyped on an ABI377 Sequencer (Genotyper 2.0; Applied Biosystems, Foster City, CA, USA), using a universal M13 sequence tagged to the forward primer (Neilan et al. 1997). The repeat sizes were confirmed by sequencing homozygous individuals. Sequencing was carried out using specific primers on an ABI 3100 capillary sequencer (Applied Biosystems) for 35 atopic asthmatic and 35 control individuals. Scoring of the identified SNPs was carried out using single-base extension reactions by SNapshot kit (Applied Biosystems) on DNA samples of nonatopic individuals (n = 94). The polymorphisms in the promoter regions and *Alu* elements were screened for transcription-factor binding sites using MatInspector 2.2 based on TRANSFAC 4.0 (Quandt et al. 1995). The haplotypes for IL4RA and STAT6 polymorphisms were estimated using Phase Ver 1.0 (Stephens et al. 2001). Linkage disequilibrium (LD; two loci) was measured using the standardized disequilibrium denoted by D′ (Ott 1999).

Results and discussion

To identify polymorphisms in *IL4RA* and *STAT6*, we sequenced 70 individuals. We identified 14 novel SNPs in our populations that were not reported elsewhere (Table 1). We also confirmed the presence of 12 known polymorphisms in these two genes, as summarized in Table 1. All the polymorphisms were found to follow Hardy-Weinberg equilibrium ($P > 0.05$). Sixty-two percent of the SNPs detected in *IL4RA* were transitions, of which 77% were identified in the coding region. Except for a novel coding SNP (a 92548-bp position C/T transition, both coding for asparagine) all SNPs were nonsynonymous (Table 1).

The high heterozygosity index (Het; 0.82) of *IL4RA R2* indicates this locus to be highly informative for genetic studies. A novel polymorphic CA repeat region was identified 2.3 kb upstream of the 5′ untranslated region (UTR) of *STAT6* (Het = 0.75). A previously reported CA repeat in

Table 1. Polymorphisms in the *IL4RA* gene and the *STAT6* gene: allele frequencies (heterozygosity)

TL4RA gene	
(CA) repeat (R1): A1>(CA)10: 0.012 to A7>(CA)20: 0.024	(0.141)
(CA) repeat (R2): A1>(CA)11: 0.004 to A11>(CA)25: 0.028	(0.842)
[a]AATTAGCTGGGCGTGGTGGC a/g	0.22 (0.351) (rs 1981551)
ATGTCCTGCGAACACCCCAC a/t	0.30 (0.426)
TCTCCAAGGCTGGAGTGCAG t/c	0.38 (0.475) Pr (99137) (rs 2057767)
TCTGCAGAGCCCACACGTGT a/g (Ile/Val)	0.43 (0.491) E3 (94272) (rs 3024558)
CCTGGGGCCCTGGGTTTCACTG c/g	0.52 (0.50) I3 (92691)
CTATGCAGTCAACATTTGGAGTGAAAA c/t (Arg)	0.94 (0.115) E4 (92548)
ACACAGATGTGGCCCACAGC c/g	0.48 (0.50) I4 (92285)
TTTGCTTCCTGGCCCCCCAC a/g	0.52 (0.51) I4 (92343)
CCCAAACTGGGAAACACAGA t/a	0.48 (0.50) I4 (92272)
GGGACCCCAGGTCCCATATGTCCAGAGA a/c	0.47 (0.499) I4 (91454) (rs 2239347)
GTGGCACAGCAGGCATTGGA g/c	0.50 (0.50) I4 (91414) (rs2239346)
GAACAAATGACAGACCAGTGTGTGGGACAG c/t	0.91 (0.141) I4 (91179)
TTGCCGACAAGTATACAATT a/	0.50 (0.50) I8 (78061)
GGCCACAGAGTGAGATCCTG t/	0.50 (0.50) I8 (77999)
TCAGCCTCCTGAGTAGCTGG a/	0.50 (0.50) I8 (77830)
CTGCTCCACCGCATGTACAAACTCC t/c (Q576R)	0.21 (0.335) E9 (76080) (rs1801275)
GTGAGACAGAGGCAGGTGGGCCTCCA c/t	0.59 (0.483) 3′-UTR (74938)
CAAATTGTCCCTGCTTTAGTCA t/c	0.45 (0.497) 3′-UTR (74693) (rs8832)
AACAAACATTTACAGACAGC t/a	0.37 (0.469) Alu Y3 (60209)
GTACCCAGCTCAAAACAACC c/a	0.37 (0.469) Alu Y3 (60703)
STAT6 gene	
(CA) repeat (R1): A1>(CA)10 (0.006) to A13>(CA)27 (0.01)	(0.750)
(CA) repeat (R3) : A1>(CA)11 (0.004) to A9>(CA)22 (0.001)	(0.650)
GAGTGACCTCAGGATAACTC a/g	0.50 (0.50) I17 (13711)
CATTTATTTATTTATTTATTTTGAG g/a	0.43 (0.50) AluY1 (5815) (rs 324012)

UTR, Untranslated region

[a]The nucleotide change contains a sequence of 20 bp penultimate to the single-nucleotide polymorphism (SNP). *Underlined* sequences denote Snapshot primers. Alleles, number of CA repeats, and allele frequencies for each (CA) can be obtained from the author on request at bghosh@cbt.res.in. Previously archived SNPs are given with their code and identifier (rs)

Table 2. Haplotypes of IL4RA and STAT6

Predicted haplotypes	Frequency
IL4RA haplotypes[a]	
1. 16 21 A G T A G T C T C G	0.025
2. 16 12 G A T G G A A T C A	0.035
3. 16 12 A G C G G T C T C G	0.050
4. 16 21 A G T G G T C T C G	0.050
5. 16 12 G A T A G A A C C A	0.030
6. 16 21 A G C G G T C T C G	0.020
7. 16 12 G A T A G A A C C G	0.070
8. 16 19 G A T A G A A T C G	0.050
9. 16 21 G A C G G T C C C A	0.045
10. 16 21 A G T A G T C C C G	0.035
11. 16 21 G A T G G A A C C A	0.035
12. 16 18 G A T A G A A C C A	0.025
13. 16 21 G A C G G A A C C A	0.045
STAT6 haplotypes	
1. 16 15	0.161
3. 16 16	0.061
4. 16 17	0.050
5. 17 15	0.033
6. 24 16	0.233
7. 24 17	0.238

[a] Haplotypes were calculated as mentioned in Subjects and Methods. The input files for Phase were generated on 182 chromosomes for IL4RA and STAT6 polymorphisms and run for 40 000 iterations. The seed values and iterations were increased by factors of two until the goodness of fit stabilized, beginning with the default values. Haplotypes occurring most frequently (at least 0.025 for IL4RA and STAT6) have been tabulated. Alleles for the haplotypes corresponded to the nucleotide in the reference sequence and occurred in genomic order

the 5′UTR was found to be polymorphic in our population (Het = 0.62) (Table 1). No coding variants of *STAT6* were detected.

The SNPs and repeats detected were used to generate haplotypes from the population data (Table 2). The SNPs for IL4RA spanned the coding region and thus may yield informative haplotypes. Because repeats are more informative in resolving ambiguous haplotypes, the data for the repeats were also used to compute the haplotypes (Table 2). Because the Phase algorithm was developed for closely linked markers, it is possible that the two repeat regions (R1 and R2 are 74 kbp and 49.3 kbp downstream of the 3′UTR, respectively) may yield inaccurate estimates of the haplotypes. However, simulations have demonstrated that this program can accurately estimate haplotypes to a range of 0.1 cM (approximately 100 kbp) (Stephens et al. 2001). Approximately 52% of the total IL4RA haplotypes and 76% of the STAT6 haplotypes are presented in Table 2. The remaining haplotypes were found to occur in lower frequencies and therefore are not presented. All these haplotypes need to be confirmed in larger family samples to show inheritance and familial segregation. LD was estimated between the markers of the *IL4RA* and *STAT6* genes and was found to be significant (Table 3). However, the repeat markers, R1 and R2, do not seem to be in LD with the SNPs in the *IL4RA* gene. This is the first report of the LD estimates over the complete coding region.

In summary, we collected data in the Indian population on the presence of polymorphisms (SNPs, simple repeats, and *AluY*) of *IL4RA* and *STAT6* spanning a total of 147 kb.

Table 3. Linkage disequilibrium (LD) estimates for pairs of markers (D′) of the *IL4RA* and *STAT6* genes

	Locus1/2	Alleles	D_
STAT6	R1/R3	24/17	0.367 (n.s)
STAT6	R1/R3	24/16	0.246 (n.s)
STAT6	R1/R3	16/15	0.365 ($p < 0.05$)
STAT6	R1/R3	24/15	0.414 ($p < 0.05$)
STAT6	R1/R3	16/16	0.361 (n.s)
IL4RA	R1/R2	16/21	0.391 (n.s)
IL4RA	R1/R2	16/12	0.866 ($p < 0.001$)
IL4RA	R2/3_-UTR (74693bp)	21/G	0.171 (n.s)
IL4RA	Q576R/I3 (92691bp)	G/C	0.663 ($p < 0.001$)
IL4RA	I3 (92691bp)/I4 (91179bp)	G/C	0.894 ($p < 0.001$)
IL4RA	E4 (92548bp)/I50V	C/G	0.742 ($p < 0.05$)
IL4RA	3_-UTR (74693bp)/I50V	G/A	0.624 ($p < 0.001$)

D′ was calculated for the region spanning the coding nucleotides

Because ethnic differences are well documented to play a major role in association studies, it is important to verify and identify polymorphisms in the particular population under study. Hence, the polymorphisms and haplotypes reported here would be valuable for genetic association studies involving the *STAT6* and *IL4RA* genes.

Acknowledgments We thank all the patients and control individuals who contributed to the study. We also thank Dr. S.K. Sharma, Dr. V. Vijayan, and Dr. V. Shandilya for helping us with sample collection. We also acknowledge the help of Mr. Dipayan Das Gupta for executing the Phase Ver1.0 program. We thank the Council for Scientific and Industrial Research and the Department of Biotechnology, Government of India, for financial assistance. R.N. acknowledges the fellowship granted by the University Grants Commission, Government of India. We also acknowledge the Functional Genomics Unit of CBT, Delhi for helping in sequencing and genotyping.

References

Barnes KC, Marsh DG (1998) The genetics and complexity of allergy and asthma. Immunol Today 19:325–332

Biedermann T, Rocken M (1999) Th1/Th2 balance in atopy. Springer Semin Immunopathol 21:295–316

Cookson W (1999) The alliance of genes and environment in asthma and allergy. Nature 402 (Suppl):B5–B11

Duetsch G, Illig T, Loesgen S, Rohde K, Klopp N, Herbon N, Gohlke H, Altmueller J, Wjst M (2002) *STAT6* as an asthma candidate gene: polymorphism-screening, association and haplotype analysis in a Caucasian sib-pair study. Hum Mol Genet 11:613–621

Hackstein H, Hecker M, Kruse S, Bohnert A, Ober C, Deichmann KA, Bein G (2001) A novel polymorphism in the 5′ promoter region of the human interleukin-4 receptor alpha-chain gene is associated with decreased soluble interleukin-4 receptor protein levels. Immunogenetics 53:264–269

Izuhara K, Yanagihara Y, Hamasaki N, Shirakawa T, Hopkin JM (2000) Atopy and the human IL-4 receptor alpha chain. J Allergy Clin Immunol 106:S65–S71

Mitsuyasu H, Yanagihara Y, Mao XQ, Gao PS, Arinobu Y, Ihara K, Takabayashi A, Hara T, Enomoto T, Sasaki S, Kawai M, Hamasaki N, Shirakawa T, Hopkin JM, Izuhara K (1999) Cutting edge: dominant effect of Ile50Val variant of the human IL-4 receptor alpha-chain in IgE synthesis. J Immunol 162:1227–1231

Nagarkatti R, Rao C-B, Rishi JP, Chetiwal R, Shandilya V, Vijayan V, Kumar R, Pemde HK, Sharma SK, Sharma S, Singh AB, Gangal SV, Ghosh B (2002) Association of *Interferon gamma* gene polymorphism with asthma in the Indian population. J Allergy Clin Immunol 110:410–412

Neilan BA, Wilton AN, Jacobs D (1997) A universal procedure for primer labelling of amplicons. Nucleic Acids Res 25:142938–142939

Nelms K, Huang H, Ryan J, Keegan A, Paul WE (1998) Interleukin-4 receptor signalling mechanisms and their biological significance. Adv Exp Med Biol 452:37–43

Ober C, Leavitt SA, Tsalenko A, Howard TD, Hoki DM, Daniel R, Newman DL, Wu X, Parry R, Lester LA, Solway J, Blumenthal M, King RA, Xu J, Meyers DA, Bleecker ER, Cox NJ (2000) Variation in the interleukin 4-receptor alpha gene confers susceptibility to asthma and atopy in ethnically diverse populations. Am J Hum Genet 66:517–526

Ott J (1999) Nonparametric methods. In: Ott J (ed) Analysis of human genetic linkage, 3rd edn. Johns Hopkins University Press, Baltimore, pp 280–284

Quandt KF, Karas KH, Wingender E, Werner T (1995) MatInd and MatInspector — New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Research 23:4878–4884

Shimoda K, van Deursen J, Sangster MY, Sarawar SR, Carson RT, Tripp RA, Chu C, Quelle FW, Nosaka T, Vignali DA, Doherty PC, Grosveld G, Paul WE, Ihle JN (1996) Lack of IL-4-induced Th2 response and IgE class switching in mice with disrupted *Stat6* gene. Nature 380:630–633

Stephens M, Smith NJ, Donnelly P (2001) A new stastical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Takabayashi A, Ihara K, Sasaki Y, Suzuki Y, Nishima S, Izuhara K, Hamasaki N, Hara T (2000) Childhood atopic asthma: positive association with a polymorphism of IL-4 receptor alpha gene but not with that of IL-4 promoter or Fc epsilon receptor I beta gene. Exp Clin Immunogenet 17:63–70

Tamura K, Arakawa H, Suzuki M, Kobayashi Y, Mochizuki H, Kato M, Tokuyama K, Morikawa A (2001) Novel dinucleotide repeat polymorphism in the first exon of the *STAT-6* gene is associated with allergic diseases. Clin Exp Allergy 31:1509–1514

Tanaka K, Sugiura H, Uehara M, Hashimoto Y, Donnelly C, Montgomery DS (2001) Lack of association between atopic eczema and the genetic variants of interleukin-4 and the interleukin-4 receptor alpha chain gene: heterogeneity of genetic backgrounds on immunoglobulin E production in atopic eczema patients. Clin Exp Allergy 31:1522–1527

Tomkinson A, Duez C, Cieslewicz G, Pratt JC, Joetham A, Shanafelt MC, Gundel R, Gelfand EW (2001) A murine IL-4 receptor antagonist that inhibits IL-4- and IL-13-induced responses prevents antigen-induced airway eosinophilia and airway hyperresponsiveness. J Immunol 166:5792–5800