Sequence variants in COL4A1 and COL4A2 genes in Ecuadorian families with keratoconus

Justyna A. Karolak,1 Karolina Kulinska,1,2 Dorota M. Nowak,1 Jose A. Pitarque,3 Andrea Molinari,3 Małgorzata Rydzanicz,4 Bassem A. Bejjani,4 Marzena Gajecka1,2

1Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2Basic Medical Sciences Program, WWAMI (Washington, Wyoming, Alaska, Montana, and Idaho), Washington State University, Spokane, WA; 3Department of Ophthalmology, Hospital Metropolitano, Quito, Ecuador; 4Signature Genomics, Spokane, WA

Purpose: Keratoconus (KTCN) is a non-inflammatory, usually bilateral disorder of the eye which results in the conical shape and the progressive thinning of the cornea. Several studies have suggested that genetic factors play a role in the etiology of the disease. Several loci were previously described as possible candidate regions for familial KTCN; however, no causative mutations in any genes have been identified for any of these loci. The purpose of this study was to evaluate role of the collagen genes collagen type IV, alpha-1 (COL4A1) and collagen type IV, alpha-2 (COL4A2) in KTCN in Ecuadorian families.

Methods: COL4A1 and COL4A2 in 15 Ecuadorian KTCN families were examined with polymerase chain reaction amplification, and direct sequencing of all exons, promoter and intron-exon junctions was performed.

Results: Screening of COL4A1 and COL4A2 revealed numerous alterations in coding and non-coding regions of both genes. We detected three missense substitutions in COL4A1: c.19G>C (Val7Leu), c.1663A>C (Thr555Pro), and c.4002A>C (Glu1334His). Five non-synonymous variants were identified in COL4A2: c.574G>T (Val192Phe), c.1550G>A (Arg517Lys), c.2048G>C (Gly683Ala), c.2102A>G (Lys701Arg), and c.2152C>T (Pro718Ser). None of the identified sequence variants completely segregated with the affected phenotype. The Glu1334His variant was possibly damaging to protein function and structure.

Conclusions: This is the first mutation screening of COL4A1 and COL4A2 genes in families with KTCN and linkage to a locus close to these genes. Analysis of COL4A1 and COL4A2 revealed no mutations indicating that other genes are involved in KTCN causation in Ecuadorian families.

Keratoconus (KTCN, OMIM 148300) is a non-inflammatory, usually bilateral disorder of the eye, characterized by progressive thinning and protrusion of the central cornea which results in altered refractive powers and loss of visual acuity [1]. The prevalence of the disease is estimated to be 1 in 2,000 individuals, and is the most common ectatic disorder of the cornea [1]. KTCN afflicts males and females in all ethnic groups [1]. Signs and symptoms depend on the stage of disease, with the first signs usually appearing in the third decade of life [1,2]. The cause of KTCN is still unknown; both genetic and environmental factors seem to play a role in its etiology. Although most cases of KTCN are isolated, an association with many syndromes, such as Down syndrome [3], Ehlers-Danlos syndrome [4], and Leber congenital amaurosis [5] has been described. Furthermore, extensive studies have shown an association between KTCN and constant eye rubbing [6], contact lens wear [7], or atopy [8]. Usually, KTCN is a sporadic disorder, but positive family history has been observed in 6%–8% of cases [1]. An autosomal dominant inheritance pattern with reduced penetrance has been suggested in 90% of patients with familial KTCN [9,10].

Genomewide linkage analyses have indicated several loci involved in the etiology of familial KTCN at 1q22.3-q23.1 (KTCN2; OMIM 608932), 3p14-q13 (KTCN3; OMIM 608586), 2p24 (KTCN4; OMIM 609271), 1p36.23–36.21, 5q14.3-q21.1, 5q21.2, 5q23-33, 8q13.1-q21.11, 9q34, 14q11.2, 14q24.3, 15q2.32, 15q22.33-q24.2, 17p13, and 20q12 [10-20]. However, no mutations in any genes at any of these loci have been associated with KTCN.

We have demonstrated an evidence of linkage to a novel locus at 13q32 [21]. Collagen type IV, alpha-1 (COL4A1; OMIM 120130) and collagen type IV, alpha-2 (COL4A2; OMIM 120090) are mapped in close proximity to that locus. The COL4A1 and COL4A2 genes are organized in a head-to-head conformation [22]. These gene pairs share a common promoter, and each gene is transcribed in opposite directions [23]. The COL4A1 gene is placed on the minus strand and consists of 52 exons, while the COL4A2 gene is on the opposite strand and consists of 48 exons. They encode two of six collagen type IV chains – α1 and α2 (1,669 and 1,712 amino acids, respectively) – forming a heterotrimeric protein molecule of collagen type IV (α1α1α2), which is found in the...
structure of the basement membrane (BM) [22,23]. Each chain contains three domains: an NH2-terminal 7S domain, a major collagenous domain with Gly-X-Y repeats (the X position is frequently occupied by proline, whereas the Y position is often occupied by 4-hydroxyproline) and a non-collagenous domain (NC1) at the COOH-terminus. Repetitions of the Gly-X-Y motif determine the formation of the triple-helical structure of collagen [22].

Collagens are the major protein components of the human cornea, and several types of collagen, including collagen type IV, have been identified [24]. Biochemical studies have revealed thinning of corneas from patients with KTCN, which may occur as a result of a reduced amount of total collagen proteins [25] and changes in collagen fibers orientation [26]. Moreover, a cornea affected by KTCN contains defects in BM and alterations in the BM composition [27]. The presence of collagen type IV in normal human cornea has remained unclear [28]. Results from expression arrays have shown an expression of COL4A1 in transplant-quality human donor corneas [29] and a downregulation of COL4A1 in keratoconus corneas [30]. Immunohistochemical studies have found collagen type IV α1/α2 chains in keratoconus corneas in large defect sites [28]. In light of these results, we recognize COL4A1 and COL4A2 as candidate genes for KTCN.

The purpose of this study was to screen COL4A1 and COL4A2 genes and determine whether sequence variants in these genes are involved in the causation of KTCN in Ecuadorian families.

METHODS

Subjects: Twenty-three individuals from family KTCN-014, 25 affected individuals from other Ecuadorian families with KTCN, and 64 Ecuadorian control subjects were included in the study. The pedigrees of these families have been described elsewhere [21]. All individuals were examined in the Hospital Metropolitano in Quito, Ecuador, undergoing a complete ophthalmic evaluation as previously described [21]. The possible consequences of the study were explained and informed consent was obtained from all family members, according to the Declaration of Helsinki. Study protocol was approved by both the Institutional Review Board at Washington State University Spokane, Spokane, WA and Poznan University of Medical Sciences (Poland).

Sequencing analyses: Oligonucleotide primers were designed to amplify all coding sequences and intron-exon junctions, promoter, and UTRs of both COL4A1 and COL4A2 (Table 1). PCR amplifications were performed using Taq DNA Polymerase (Fermentas Inc., Glen Burnie, MD). PCR products were purified with ExoSAP-IT® (USB Corporation, Cleveland, OH) or Montage® PCR Filter Units (Millipore, Jaffrey, NH) and sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Inc. [ABI], Foster City, CA). Sequencing was visualized on an ABI PRISM® 3100 Genetic Analyzer (ABI) and a 3730xl DNA Analyzer (ABI). The DNA sequences of study subjects were compared with the reference sequences of COL4A1 and COL4A2 (GRCh37/hg19, GenBank accession numbers for the mRNA NM_001845.4 and NM_001846.2, respectively) using Sequencher® 4.1.4. Software (Gene Codes Corporation, Ann Arbor, MI).

Haplotypes analysis: PEDSTATS [31] was used to verify the structure of KTCN-014 family and identify potential Mendelian inconsistencies in the inheritance of single nucleotide polymorphisms (SNPs) in COL4A1 and COL4A2. For that region, to determine the full haplotypes inherited along with the substitutions occurring in affected individuals, a reconstruction of observed sequence variants was prepared using SimWalk2 [32,33]. Allele frequencies were set as equal. The location of genetic markers was determined on the basis of the Rutgers combined linkage-physical map of the human genome [34], either directly or by interpolation. Haplotype was generated with HaploPainter [35].

Statistical analysis for Gln1334His substitution: The difference in distribution of Gln1334His substitution between affected and unaffected individuals in family KTCN-014 was analyzed by Fisher's Exact Test for Count Data. Similarly, 25 affected individuals from the remaining KTCN families versus 64 Ecuadorian control individuals were compared using Fisher's Exact Test. The difference between the examined groups was considered significant if the value of probability (p) did not exceed 0.05.

Prediction of effect of amino acid substitutions on protein function: The potential impact of amino acid substitutions on the COL4A1 and COL4A2 proteins was examined using PolyPhen, SIFT, PMUT, PANTHER, and SNAP tools. The PolyPhen tool predicts which missense substitution affects the structure and function of protein, and uses Position-Specific Independent Counts software to assign profile scores. These scores are the likelihood of the occurrence of a given amino acid at a specific position, compared to the likelihood of this amino acid occurring at any position (background frequency) [36].

The SIFT analytic tool, on the basis of gene sequences homology, evaluates conserved positions, and calculates a score for the amino acid change at a particular position. A score of <0.05 is considered as pathogenic and has a phenotypic effect on protein structure [37]. The PMUT calculates the pathological significance of non-synonymous amino acid substitution using neural networks (NN). NN output >0.5 is considered to be deleterious [38]. PANTHER estimates the likelihood of a particular amino acid’s change affecting protein function. On the basis of an alignment of evolutionarily related proteins, it generates the substitution Position-Specific Evolutionary Conservation (subPSEC). The subPSEC could achieve values from 0 (neutral) to about −10 (most likely to be deleterious). The
| Name       | Forward                     | Reverse                     | Annealing Temperature (°C) | Amplicon Size (bp) |
|------------|-----------------------------|-----------------------------|---------------------------|--------------------|
| COL4A1.1   | CACCCCTCCCCCTTTCTACTC       | GCCCAGAGAATGCACCTG          | 59                        | 837                |
| COL4A1.2   | TTGGGCTGAGTAACACTTGG        | GCCCTTTTGAGCTTCATTG         | 58                        | 459                |
| COL4A1.3-4 | GGCAACAGAAATGAGACTCC        | TGTGGAGCTGGAGAGGAGAGAT      | 66                        | 477                |
| COL4A1.5-6 | GGGACAGCTGCTTGGTGTG         | ACAAAGCTGTACTGAGGTA         | 60                        | 698                |
| COL4A1.7-8 | CACACTGTGGAGGAAGGTAGG       | TGTCCAGTGTCTGACAG           | 58                        | 578                |
| COL4A1.9-10| CTTTGTCTCCCTGGCCTC          | TCACTCTCCCTCTCCCACAG        | 58                        | 691                |
| COL4A1.11  | GGAATGATGTTGGTATGGTG        | GACTAAGGGGTGGAGGAAGAGG      | 58                        | 451                |
| COL4A1.12  | GGAGAAGCCATGATGGCTGA        | GACATTGATCCAAGGTTGGG        | 58                        | 239                |
| COL4A1.13  | GCAGAGACAGATGAGTACAG        | GGAGGCTGCTTATGTAAGGC        | 58                        | 393                |
| COL4A1.14-15| CTTGCTCCCTGCTTACATT        | GTCCTCAGAGCCTCTCTCT         | 60                        | 505                |
| COL4A1.16-17| TTATGGGAGAGGAGATGGTTG       | AAATGGTCTCTGATGACTG         | 60                        | 725                |
| COL4A1.18  | GATGGGACAAAGTACCTGGG        | CATCTGCTTCTCTCTCCTC         | 60                        | 459                |
| COL4A1.19  | GCTACCATTTGCTGCTATCTAC      | AATAGAAGAGCTGAGGAGAGAG      | 62                        | 447                |
| COL4A1.20  | GTCAACACAGGCTTACAGAG        | GCCAGGAGAGACATAGGGGT        | 60                        | 486                |
| COL4A1.21  | CAGTGTGCTTGTGTGGATG         | ATGCCAGAGTCTCAGAGG          | 60                        | 516                |
| COL4A1.22  | TGGGGCTGAGTAAGGTCAAGG       | GAGAAGGGCGAAAACCTCTCA       | 60                        | 304                |
| COL4A1.23  | TCCACACCATTGAGCAGAGAG       | GCAACACACACCAAAGGCGA        | 60                        | 431                |
| COL4A1.24  | GTGGCTCTTGGGATTTTATG        | ATTTGGGTCTGTTGGAATC         | 60                        | 718                |
| COL4A1.25  | GTGCCAAAGCCATCTACATA        | GTTGCACAGGTTGCCCAATG         | 60                       | 488                |
| COL4A1.26  | CCTGGGAGGAGGATAGAAGT         | GAAGGAGGAGCACAAGAGG         | 62                        | 488                |
| COL4A1.27-28| AAATGGGAGAGGAGATGGTTG       | TCTTCCACACACACCTTAC         | 56                        | 636                |
| COL4A1.29  | AGGGTGCTGAGGAAGACAGCA       | GCTAGAAGCTGAGAAAACCATC      | 60                        | 678                |
| COL4A1.30  | GCTGGGCTGAGGAAGACAGCA       | GGCTTACAGATTTGCTACG         | 64                        | 315                |
| COL4A1.31  | CACAGCCCTACAGCAGTTA         | CAGTGTGCGAGGAGAAACATC       | 61                        | 483                |
| COL4A1.32-33| CTTACAGTTTCTCCAGTGTG         | GCCCTTCTGAGTGTCTTCA         | 60                        | 653                |
| COL4A1.34  | CTGATCCAGCCCTGAGGTA         | TATGGGAGGACCGATAACCC         | 60                        | 411                |
| COL4A1.35-36| TGGCCCTTTCTGCTGGTAAT        | AATGGCAGGACCTCCAGTAC         | 64                        | 594                |
| COL4A1.37  | GGGAGGGATTACCTCTGTTGTA      | TCCCTGCTGGTGTAATGGCTCA      | 58                        | 364                |
| COL4A1.38-39| TGGCCGTGAGGAGGAGAGAT        | TGAAGATGGGAGAGCAGAC         | 61                        | 641                |
| COL4A1.40  | GACCTCAGGAAAACAGGGT         | GATGTGCGAGGAGTGCAG          | 60                        | 359                |
| COL4A1.41  | TGGTGCTGTCGTGAGCTGAGG       | CATGTGCTGCTGAGCAGC          | 60                        | 447                |
| COL4A1.42  | TAAAGAAGAGGGAGGAGATGG        | TCTTCAACGAGACGCAAAGG         | 62                        | 676                |
| COL4A1.43  | CTCGCGCCTGAGTTTCTGGA        | TAGTGCGAGGATGAGGGT          | 60                        | 435                |
| COL4A1.44  | CCACAGGCCATCTTGGTGT         | TACAAATGGGTGGCTGCAC         | 60                        | 376                |
| COL4A1.45  | GGACCAAGAACAATGCTCTCA       | GAGCCTTGGGAAGTCTGTA         | 60                        | 790                |
| COL4A1.46  | CGAAGATGGTGTGGAGAGT         | TCTCTGGGTTTCTCTGTG          | 59                        | 590                |
| COL4A1.47  | AAGCCAAAGAACAGGGGGAAGT      | GAGCTGGCTTTCACACAAATC       | 60                        | 591                |
| COL4A1.48  | TGGAGAAGGGATGAGCAGG         | GCGAAGTGTCTACTGCTAC         | 60                        | 516                |
| COL4A1.49  | GTGTGCTGTAAGACATGCTCC       | GCACACGCTTGAACCTTTCATTTA    | 60                        | 650                |
| COL4A1.50  | AAAACCAAGGGGAGGATG         | TAAAGCGAGACGAGATAAGGAGA     | 59                        | 407                |
| COL4A1.51  | GGAAGCAGCCATTAGAGGAT       | AAATGCCTCTGGTCTACTGCG       | 60                        | 573                |
| Name       | Forward | Reverse |
|------------|---------|---------|
| COL4A1.52.1 | TACCGTGTGAGGCCCTGATG | ACCTCCTAGCAACCTTTTGGT |
| COL4A1.52.2 | GAAACCACAGGGGTCTAGG | CCGAATGTGCTTTAGCTGTA |
| COL4A1.52.3 | CCTGGCTTGAAMAACAGCTC | ATCCACCCCCAGCTGTCAG |
| COL4A2.1   | TCTGGGQAAGACTGAGATAC | AGACAAAGGAGGATGAGAGA |
| COL4A2.2   | GCTCTGGAAGGGCCTATGC | GGAGAAGGGAGAGAAGAGA |
| COL4A2.3   | CTCATCTCAAGCTTACCTC | ACATCTTCTCCCTCAGTCA |
| COL4A2.4   | ATTCAGGGTGGAGAGAC | CGGCCATCTAGTGTGGT |
| COL4A2.5–6 | TCTTTTTATCCCAACAGCTC | TCCACGTTTTTTATGTC |
| COL4A2.7   | AGACAGAAGAAACCCCGACA | TCTGGGGTCAACACAGAT |
| COL4A2.8   | TGCAATAAACCCCTACAGC | AACAGATCGCCCTACAGGAC |
| COL4A2.9–10| AGGTGCGTATGGGCTGATCT | TAACTGCGAGAGCCTGAGT |
| COL4A2.11  | GCACTGAGAAAACCTCCATGC | ACATTGGCCCTCCCCAACA |
| COL4A2.12  | TGCAATCTCAGCCTCCTCAGA | CCCAGTCTAGTGTGGT |
| COL4A2.13  | GGAAACAACCCACAGAAAC | GCCAAATGAGAGCAGGAT |
| COL4A2.14  | GTAAAACATCTGCCTGGAAG | CTATGGAACAGGGGAGT |
| COL4A2.15  | GTCACTGCTGTCCTCAGA | CCCAGTCTAGTGTGGT |
| COL4A2.16  | ATTAATTTCCATCCACCTC | GGAGGACCGGTTAATG |
| COL4A2.17  | CGCAGTGTCCTCAACACACCA | CGTGAAGGCTCAGATTG |
| COL4A2.18  | AGCAAGCTGCTGCAATTTTC | CTGTAAGGGGTTCCTACG |
| COL4A2.19  | TTCTGAGCTGTGGAGAGCTC | AGCTGCTAGGAGGAT |
| COL4A2.20  | ACCCATCGGATTTAGTGGC | AGATCGCCCTACTقات |
| COL4A2.21  | CCTGCATCTGCTGTCTC | AAGTTGCTCTCCCTTCAC |
| COL4A2.22  | CTCCTGGTAGATGTGCTCAGT | AAGATGCTGCTTCTTCAT |
| COL4A2.23–24| ATGCAAGAAGAATCTCCCTTGG | ATGAGCGCTCCCTCTATG |
| COL4A2.25  | TGCCACTAGTCTGTTCCTCA | ACAGAGAGGAGGATGTTT |
| COL4A2.26  | AACATCGACCTCCTGTCTG | TTCTGACAGAGGGGAGT |
| COL4A2.27  | CGAAGATGTGAGGCCTTCTT | GCAAAGCAAGGATGGTG |
| COL4A2.28  | TAAAGCCTGGAGGCTGTTT | CGAAACACCTGTCTCCTT |
| COL4A2.29  | GCCAAGGTGTAGTGGTTCCTA | TGGCAACAGAAAAACAGT |
| COL4A2.30  | GAAATAGCAAGGGCAGGAGG | CGAAGAGATAGGCCGAT|
| COL4A2.31  | CACGAGCTCAACTCCCTAGAT | AGCAGAGGAGGATGTTT |
| COL4A2.32  | TGTCCTCTCTGCTTTCCTT | TGTTGACAGGAGATAGA |
| COL4A2.33  | TGTCCTCTCTCCACCTGTC | AACCAGGTTACACAGG |
| COL4A2.34  | AAGCAGCTAGAGCACAAAA | ACATCTGCAATGTCCTAA |
| COL4A2.35  | GCTATAGAAACGCCCTATG | ACAGAGCTTCCCTACGAG |
| COL4A2.36  | GGAGTCCCAATACTCAGAGC | GACCCTCTGGCTTTCTGAG |
| COL4A2.37  | CCCATGCTCTTCTCTCATTG | ATGGCTCTCTCCATCCCTGT |
| COL4A2.38  | CTGTCGCTCTTTTCTGTTT | CTTGTCGCTATTTTCG |
| COL4A2.39  | GTGCTGCTCCCCACATGAAA | AGTCCAATTCCCTACG |
| COL4A2.40  | ATGCGGCTCTGATCCTCTT | AAAACAGCTTCTCTCCG |
| COL4A2.41  | CCCACCACTAGAGATGTTT | ATGACACAGGGAGGCTCAT |
| Name       | Forward                           | Reverse                          | Annealing Temperature (°C) | Amplicon Size (bp) |
|------------|-----------------------------------|----------------------------------|---------------------------|--------------------|
| COL4A2.42–43 | AGTCATTCCATGCCACAGAC             | TAAGCTCTCCATTCCCCAAG             | 60                        | 666                |
| COL4A2.44–45 | CCCGTTAGTGCTGGCTCAT              | AGGTGTCTTGCTGGCATAG             | 60                        | 744                |
| COL4A2.46   | GAAACTGCCCTGCACCTCT             | TAGATGGACCTTTCCGTAGC             | 60                        | 664                |
| COL4A2.47   | CACTCCTGCTGGATCCTAACT           | CCACTACGCTTTGCTGCA GTG G        | 60                        | 675                |
| COL4A2.48.1 | GGATGCCTCATGTCCGTATTT            | TACATGGGTGTGTCGGA G             | 60                        | 689                |
| COL4A2.48.2 | CATCCAGCAGCAGCACTTAAAG          | AGGTCTCCACTTCTGCTG A             | 59                        | 530                |
| COL4A2.48.3 | CCTGCTTCTACGCAAATGT             | CTGTTCTTTTCTGCTTGT              | 60                        | 573                |

In the table, Amplicon Size represents length of the PCR product in base pairs (bp) and Annealing Temperature represents the annealing temperature of the primers used for PCR amplifications.
value −3 is the cutoff point for functional significance, and corresponds to a \( P_{\text{determin}} \) of 0.5. If the substitution occurs at a position not appearing in the multiple sequence alignment, a subPSEC score cannot be calculated and change is not likely to be pathogenic [39,40].

The SNAP tool predicts the functional consequences of exchanging amino acids using evolutionary conservation and structure/function relationships. The SNAP output shows prediction neutral or non-neutral, and the expected accuracy [41].

RESULTS

Forty eight members of 15 Ecuadorian families and 64 Ecuadorian control subjects were included in the study. Twenty-three individuals from family KTCN-014, two affected individuals from each of the families KTCN-011, 015, 019, 020, 021, 024, 025, 030, 031, 034, and 035, and one patient from each of KTCN-05, 013, and 017 were examined.

**COL4A1 and COL4A2 sequence analyses:** Screening of **COL4A1** (NM_001845.4) coding regions revealed 12 sequence variants, three of which were amino acid substitutions: c.19G>C (Val7Leu), c.1663A>C (Thr555Pro), and c.4002A>C (Gln1334His). We identified one novel synonymous change, c.3693G>A (Thr1213Thr), and eight previously reported sequence variants: c.432T>A (Ala1442Ala), c.1257T>C (Pro419Pro), c.1815T>C (Pro606Pro), c.2130T>A (Pro710Pro), c.3183G>A (Gly1061Gly), c.3189A>T (Arg1063Arg), c.4470C>T (Ala1490Ala), and c.4800C>T (Ser1600Ser). In the 5′ untranslated region (5′ UTR), one novel sequence variant, c.84+124T>A, was identified. In the 3′ untranslated region (3′ UTR), two previously reported variants, c.*587C>A and c.*975A>C, were detected.

Sequencing analyses of **COL4A2** (NM_001846.2) coding regions revealed 13 previously reported sequence variants, including five non-synonymous substitutions: c.574G>T (Val191Phe), c.1550G>A (Arg517Lys), c.2048G>C (Gly683Asp), c.2102A>G (Lys701Arg), and c.2152C>T (Pro718Ser), and eight synonymous substitutions: c.297G>A (Thr99Thr), c.1083C>T (Pro361Pro), c.1095G>A (Pro365Pro), c.1179C>T (Ile393Ile), c.1488G>A (Pro496Pro), c.4089G>A (Ala1363Ala), c.4290T>C (Phe1430Phe), c.4515A>G (Pro1505Pro). In the 5′ UTR, five known nucleotide changes, c.-227A>C, c.-223G>G, c.-215C>T, c.-203T>C, and c.-133A>G, were identified. In the 3′ UTR, eight previously reported sequence variants, c.*767C>T, c.*101*102del2, c.*417C>G, c.*541C>T, c.*557A>G, c.*650T>C, c.*663T>C, and c.*727G>C were detected.

Screening of exon/intron junctions in **COL4A1** and **COL4A2** revealed numerous sequence variants in the surrounding non-coding sequences, 71 and 86, respectively, including single nucleotide changes, insertions, and deletions. All screening results are summarized in Table 2.

The sequencing of the genomic region containing the common promoter of **COL4A1** and **COL4A2** revealed no sequence changes.

**Statistical analysis and in silico predictions:** PolyPhen analyses of non-synonymous changes in **COL4A1** and **COL4A2** predicted that only the Gln1334His variant in **COL4A1** was possibly damaging for protein function and structure. The multiple sequence alignment of **COL4A1** orthologs shows that the amino acid glutamine at position 1,334 is conserved throughout the analyzed species (Figure 1). Gln1334His substitution was observed more frequently in patients than in healthy individuals in family KTCN-014 (p=0.056). There was no difference in the c.4002A>C allele distribution between the analyzed affected individuals from the remaining KTCN families and the Ecuadorian control subjects (p=0.17).

The SIFT, PMUT, PANTHER, and SNAP analyses defined all missense amino acid substitutions in **COL4A1** and **COL4A2** as neutral/tolerated and lacking any effect on protein function. All prediction results are summarized in Table 3.

**Haplotype reconstruction:** Haplotypes of sequence variants observed in family KTCN-014 are shown in Figure 2. The coding sequence variants in **COL4A1** are surrounded by markers rs13260 and col4a1_snpl. Exons of **COL4A2** are localized between rs35466678 and rs422733.

KTCN-014 consists of two family branches. Distinct haplotypes in the branches were identified (Figure 2). In the first one, initiated by parents KTCN-93 and KTCN-01, six subjects with KTCN had the same haplotype in the **COL4A1** region, extending from rs13260 to col4a1_snpl. Three unaffected individuals, KTCN-13, KTCN-14, and KTCN-22, share that part of the haplotype with their affected relatives. One of four variants in this region, rs3742207, causes a change in the protein sequence, replacing Gln in position 1334 with His (Gln1334His). That haplotype region, from rs13260 to col4a1_snpl, represents a short fragment of the haplotype which covers the whole **COL4A1** and **COL4A2** sequence in KTCN-03, KTCN-05, KTCN-06, and KTCN-14. In addition, individuals KTCN-07, KTCN-09, KTCN-13, KTCN-22, and KTCN-23 share the rs874203-rs422733 region (Figure 2 – pink bars). For markers rs13260-col4a1_snpl, a different haplotype was observed in the second family branch, initiated by parents KTCN-92 and KTCN-16. This haplotype covered the entire length of the analyzed region, and was identified in all affected individuals and KTCN-21, whose phenotype was unknown. Subject KTCN-17 had the same allele pattern for markers rs13260-col4a1_snpl, as individuals from the first branch of the family. However, in this case, analysis indicated that these markers are inherited from KTCN-92, who is unrelated to KTCN-93 and KTCN-01.

**DISCUSSION**

To our knowledge, this is the first report describing complete sequence analysis of the coding regions and the exon-intron
### Table 2. Sequence variants found in COL4A1 and COL4A2 genes.

| Exon | dbSNP refID | Chromosome Position | Allele Change | Residue Change | Affected KTCN-014 (n=10) | Unaffected KTCN-014 (n=11) | Unknown KTCN-014 (n=2) | All KTCN-014 (n=23) | Other KTCN families affected (n=25) | All (n=48) |
|------|-------------|---------------------|---------------|---------------|--------------------------|---------------------------|-----------------------|-------------------|----------------------------------|-------------|
| 1    | rs9515185   | 11095356             | c.19G>C       | Val7Leu       | 10 100                   | 11 100.0                  | 2 100                  | 23 100.0          | 25 100.0                        | 48 100.0    |
| 2    | rs1125665   | 11095290             | c.85–196C>T   | 1 10          | 0 0                      | 2 18.2                    | 1 50                   | 3 13.0             | 5 20                             | 8 16.7      |
| 3    | rs12475106  | 11095150             | c.85–69C>T    | 1 10          | 0 0                      | 4 17.4                    | 3 14                   | 8 16.7             |                                  |             |
| 4    | rs9216510   | 110862625            | c.325+123G>A  | - 10          | 0 0                      | 1 50                      | 0 0                   | 3 13.0             | 5 20                             | 8 16.7      |
| 5    | rs737328    | 110866065            | c.279+64G>A   | 1 50          | 5 45.5                   | 1 50                      | 11 47.8               | 10 40              | 21 43.8                          |             |
| 6    | rs522625    | 110864225            | c.432T>A      | Ala144Ala     | 7 70                      | 8 72.7                    | 2 100                  | 17 73.9            | 14 56                            | 31 64.6     |
| 7    |             | 110863956            | c.468+5_468+9del5 | - 3 30 | 2 18.2 | 0 0 | 5 | 21.7 | 4 16 | 9 | 18.8 |
| Exon | dbSNP refID | Chromosome | Allele Change | Residue Change | Affected KTCN-014 (n=10) | Unaffected KTCN-014 (n=11) | Unknown KTCN-014 (n=2) | All KTCN-014 (n=23) | Other KTCN families affected (n=25) | All (n=48) |
|------|-------------|------------|---------------|----------------|--------------------------|--------------------------|------------------------|-----------------|----------------------------------|-----------|
| 25   | rs556174    | 110839550  | c.1663A>C     | Thr555Pro      | 10 100 11 100.0 2 100 23 100.0 25 100 48 100.0 |
| 26   | rs9526138   | 110839428  | c.1728+57T>C  | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 27   | rs81749673  | 110838703  | c.1831T>C     | Pro605Pro      | 3 30 4 36.4 1 50 8 34.8 4 16 12 25.0 |
| 28   | rs2503080   | 110838646  | c.1991–16G>A  | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 29   | rs61749897  | 110838519  | c.2095+145G>T | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 30   | rs874204    | 110837502  | c.2130G>A     | Pro710Pro      | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 31   | rs16975492  | 110833702  | c.2194–69C>T  | -              | 3 30 4 36.4 1 50 0 0.0 0 0 6 34.8 |
| 32   | rs1319599   | 110831866  | c.2194–69C>T  | -              | 3 30 4 36.4 1 50 0 0.0 0 0 6 34.8 |
| 33   | rs903053    | 110831451  | c.2345–68A>G  | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 34   | rs350589    | 110830612  | c.2626–34T>C  | -              | 0 0 0 0 0 0 0 0 0 0 2 8.3 |
| 35   | rs16975491  | 110833564  | c.2893+57T>C  | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 36   | rs10492497  | 110831837  | c.2893+57T>C  | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 37   | rs874203    | 110837540  | c.3183G>A     | Gly1061Gly     | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 38   | rs874201    | 110837540  | c.3389A>T     | Arg1063Arg     | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 39   | rs874204    | 110837540  | c.3439G>A     | Thr1231Thr     | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 40   | rs61749897  | 110838519  | c.3557–90C>T  | -              | 3 30 4 36.4 1 50 8 34.8 6 24 14 29.2 |
| 41   | rs874203    | 110837540  | c.3590+136C>T | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 42   | rs7151998   | 110825264  | c.3596–147C>A | -              | 1 10 1 9.1 0 0 2 8.3 4 16 8 12.5 |
| 43   | rs2289799   | 110828497  | c.3596–147C>A | -              | 1 10 1 9.1 0 0 2 8.3 4 16 8 12.5 |
| 44   | rs2275845   | 110822943  | c.3693G>A     | Thr1231Thr     | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 45   | rs899865    | 110819586  | c.3771–90C>T  | -              | 7 70 7 63.6 1 50 15 65.2 2 100 23 100.0 |
| 46   | rs652572    | 110819457  | c.3949+48T>C  | -              | 2 20 4 36.4 1 50 8 34.8 6 24 14 29.2 |
| 47   | rs1213026   | 110819362  | c.4449–143T>C | -              | 9 90 10 90.9 2 100 21 91.3 24 96 45 93.8 |
| 48   | rs1133219   | 110817097  | c.4470C>T     | Ala1490Ala     | 3 30 5 54.5 2 100 14 52.2 12 66 26 54.2 |
| 49   | rs2275843   | 110813573  | c.4590+137C<T | -              | 6 20 4 36.4 1 50 8 34.8 6 24 14 29.2 |
| 50   | rs1319599   | 110813531  | c.4620+86A>G  | -              | 6 20 4 36.4 1 50 8 34.8 6 24 14 29.2 |
| 51   | rs2275842   | 110813531  | c.4620+86A>G  | -              | 2 10 1 9.1 0 0 2 8.3 4 16 8 12.5 |
| 52   | rs2275842   | 110813531  | c.4620+86A>G  | -              | 3 30 4 36.4 1 50 8 34.8 6 24 14 29.2 |
| 53   | rs605078    | 110804899  | c.4800C>T     | Ser1600Ser     | 2 20 4 36.4 1 50 8 34.8 6 24 14 29.2 |
| 54   | rs13260     | 110802123  | c.587C>A      | -              | 1 10 1 9.1 0 0 2 8.3 4 16 8 12.5 |
| 55   | rs28362515  | 110801735  | c.975A>C      | -              | 1 10 1 9.1 0 0 2 8.3 4 16 8 12.5 |
| Exon | dbSNP refID | Chromosome | Allele Change | Residue Change | Affected KTCN-014 (n=10) | Unaffected KTCN-014 (n=11) | Unknown KTCN-014 (n=2) | All KTCN-014 (n=23) | Other KTCN families affected (n=25) | All (n=48) |
|------|-------------|------------|---------------|----------------|--------------------------|--------------------------|------------------------|-----------------|-----------------------------------|-----------|
| Ex5  | rs74967960   | 110977274  | c.297G>A      | Thr99Thr       | 10 100 11 100.0         | 2 100 23 100.0          | 22 88 45 93.8         |                 |                                   |           |
| Ex9  | rs7334986    | 110977696  | c.361–205G>A  | Val192Phe      | 10 100 11 100.0         | 2 100 23 100.0          | 22 88 45 93.8         |                 |                                   |           |
| Ex17 | rs4103       | 110988226  | c.1008C>T     | Pro336Pro      | 10 100 11 100.0         | 2 100 23 100.0          | 22 88 45 93.8         |                 |                                   |           |
| Ex19 | rs7498905    | 110990005  | c.1339–52G>A  | Val1191Pro     | 10 100 11 100.0         | 2 100 23 100.0          | 22 88 45 93.8         |                 |                                   |           |
| Ex22 | rs7990214    | 110997585  | c.1480G>A     | Pro496Pro      | 10 100 11 100.0         | 2 100 23 100.0          | 22 88 45 93.8         |                 |                                   |           |

Molecular Vision 2011; 17:827-843 http://www.molvis.org/molvis/v17/a94 © 2011 Molecular Vision
### TABLE 2. CONTINUED.

| Exon | dSNP | rsno | Allele Change | dbSNP refID | Chromosome | Allele | Residue Change | no. (%) | no. (%) | no. (%) | no. (%) | no. (%) |
|------|------|------|---------------|-------------|------------|--------|----------------|--------|--------|--------|--------|--------|
| Ex27 | rs9559818 | 111130226 | c.2426–124G>A | 8 80 7 63.6 | 1 50 16 69.6 | 21 84 37 77.1 |
| Ex28 | rs9521803 | 111132556 | c.2588–11C>T | 8 80 6 54.5 | 2 100 16 69.6 | 21 84 37 77.1 |
| Ex28 | rs7790301 | 111132763 | c.2588–154C>T | 4 40 2 18.2 | 0 0 5 25.0 | 8 32 12 25.0 |
| Ex31 | rs58124222 | 111132947 | c.2758+210G>A | 0 0 3 27.3 | 0 0 3 13.0 | 8 32 11 22.9 |
| Ex32 | rs3803229 | 111134780 | c.2759–83G>A | 6 60 7 63.6 | 1 50 16 69.6 | 21 84 37 77.1 |
| Ex32 | rs9559817 | 111134858 | c.2759–5T>C | 0 0 3 27.3 | 0 0 3 13.0 | 8 32 11 22.9 |
| Ex32 | rs41315048 | 111137975 | c.3026–27G>T | 1 10 3 27.3 | 0 0 4 16.7 | 6 24 11 22.9 |
| Ex32 | rs2296851 | 111138255 | c.3207+72G>A | 2 20 2 18.2 | 0 0 4 16.7 | 6 24 11 22.9 |
| Ex32 | rs452020 | 111144103 | c.3455–315T>C | 10 100 11 100.0 | 2 100 10 90.9 | 25 100 50 100.0 |
| Ex32 | rs403839 | 111144321 | c.3455–96G>A | 2 20 2 18.2 | 0 0 4 16.7 | 6 24 11 22.9 |
| Ex32 | rs421177 | 111144565 | c.3562+41C>T | 1 10 3 27.3 | 0 0 4 16.7 | 6 24 11 22.9 |
| Ex32 | rs57003582 | 111145456:111145486 | 70del30 | 1 10 3 27.3 | 0 0 4 16.7 | 6 24 11 22.9 |
| Ex32 | rs227454 | 111114217 | c.3799+4G>A | 2 20 2 18.2 | 0 0 4 16.7 | 6 24 11 22.9 |
| Ex32 | rs3803237 | 111114675 | c.3803237+3G>A | 10 100 11 100.0 | 2 100 10 90.9 | 25 100 50 100.0 |

© 2011 Molecular Vision
### Table 2. Continued.

| Exon | dbSNP ref ID | Chromosome | Allele Change | Residue Change | Affected KTCN-014 (n=10) | Unaffected KTCN-014 (n=11) | Unknown KTCN-014 (n=2) | All KTCN-014 (n=23) | Other KTCN families affected (n=25) | All (n=48) |
|------|--------------|------------|----------------|----------------|------------------------|-------------------------|-----------------------|------------------------|------------------------------------|-----------|
|      |              |            |                |                | no. | %     | no. | %     | no. | %     | no. | %     | no. | %     | no. | %     | no. | %     | no. | %     | no. | %     | no. | %     |
| Ex43 | rs378601     | 1111153934 | c.3761–81G>A   | 10             | 100      | 11   | 100.0 | 2   | 100      | 23   | 100.0 | 25   | 100      |
|      | rs380222     | 1111154159 | c.3777+28C>T   | 8              | 80       | 8    | 72.7  | 2   | 100      | 18   | 78.3   | 21   | 84       | 39   | 81.3   |
|      | rs2281968    | 1111154160 | c.3777+29G>A   | 5              | 50       | 5    | 62.7  | 1   | 50       | 14   | 60.9   | 18   | 72       | 32   | 66.7   |
|      | rs4773198    | 1111155711 | c.4040-19C>T   | 5              | 50       | 5    | 56.3  | 1   | 50       | 12   | 52.2   | 8    | 32       | 20   | 41.7   |
|      | rs4773199    | 1111155779 | c.4089G>A     | 5              | 50       | 4    | 32.2  | 1   | 50       | 10   | 45.3   | 8    | 32       | 18   | 37.5   |
| rs69301400 | 1111156155  | c.4139-41G>A  | 5              | 50       | 6    | 45.5  | 1   | 50       | 12   | 52.2   | 8    | 32       | 20   | 41.7   |
|      | rs414881     | 1111156411 | c.4258+71G>A   | 7              | 70       | 8    | 57.7  | 1   | 50       | 16   | 69.6   | 21   | 84       | 37   | 77.1   |
|      | Ex46         | rs4771683   | c.4290T>C     | 10             | 100      | 11   | 100.0 | 2   | 100      | 23   | 100.0 | 25   | 100      |
|      | rs445346     | 1111156499 | Phc1-430Phe    | 10             | 100      | 11   | 100.0 | 2   | 100      | 23   | 100.0 | 25   | 100      |
|      | rs2479426    | 1111164198 | c.4882-83T>C   | 7              | 70       | 6    | 65.5  | 2   | 100      | 15   | 65.2   | 20   | 80       | 35   | 72.9   |
|      | rs422733     | 1111164641 | c.76T>C       | 7              | 70       | 6    | 54.5  | 2   | 100      | 15   | 65.2   | 20   | 80       | 35   | 72.9   |
|      | rs3074455    | 1111164639, | c.*101_102del2 | 7              | 70       | 6    | 54.5  | 2   | 100      | 15   | 65.2   | 20   | 80       | 35   | 72.9   |
|      | rs10599      | 111114955  | c.541C>G      | 8              | 80       | 11   | 100.0 | 2   | 100      | 21   | 95.3   | 25   | 100      |
|      | rs1049906    | 1111154179 | c.*541C>T     | 4              | 40       | 8    | 72.7  | 1   | 50       | 13   | 56.5   | 18   | 72       | 31   | 64.6   |
|      | rs1049931    | 1111165095 | c.*557A>G     | 6              | 60       | 9    | 81.8  | 1   | 50       | 16   | 69.6   | 20   | 80       | 36   | 75.0   |
|      | rs1049777    | 1111165188 | c.*691T>C     | 6              | 60       | 9    | 81.8  | 1   | 50       | 16   | 69.6   | 19   | 76       | 35   | 72.9   |
|      | rs7711       | 1111165201 | c.*663C>T     | 6              | 60       | 9    | 81.8  | 1   | 50       | 16   | 69.6   | 19   | 76       | 35   | 72.9   |
|      | rs15457      | 1111165265 | c.*727G>C     | 4              | 40       | 8    | 72.7  | 1   | 50       | 13   | 54.6   | 17   | 68       | 30   | 62.5   |

**dbSNP ref ID:** identity numbers of observed sequence variants; chromosome position (NCBI build 37.1).
| Gene   | Sequence variant | PolyPhen | SIFT | PMUT | PANTHER | SNAP |
|--------|-----------------|----------|------|------|---------|------|
|        |                 | PSIC score | Prediction | Score | Prediction | NN | Prediction | subPSEC | Pdeleterious | Expected Accuracy | Prediction |
| COL4A1 | Val7Leu         | N/A       | benign | 1    | tolerated | 0.2367 | neutral | - | - | 92% neutral |
|        | Thr555Pro       | N/A       | benign | 0.65 | tolerated | 0.0250 | neutral | -0.52603 | 0.0777 | 94% neutral |
|        | Gln1334His      | 1.66      | possibly damaging | 0.12 | tolerated | 0.1039 | neutral | -1.0433 | 0.12382 | 69% neutral |
| COL4A2 | Val192Phe       | 1.13      | benign | 64   | tolerated | 0.1921 | neutral | - | - | 78% neutral |
|        | Arg517Lys       | 0.1       | benign | 0.96 | tolerated | 0.0861 | neutral | - | - | 92% neutral |
|        | Gly683Ala       | N/A       | benign | 0.96 | tolerated | 0.4841 | neutral | - | - | 85% neutral |
|        | Lys701Arg       | N/A       | benign | 0.97 | tolerated | 0.0166 | neutral | - | - | 89% neutral |
|        | Pro718Ser       | N/A       | benign | 0.98 | tolerated | 0.2039 | neutral | - | - | 89% neutral |

The PolyPhen tool predicts which missense substitution affects the structure and function of protein, and uses Position-Specific Independent Counts software to assign profile scores. The SIFT tool evaluates conserved positions, and calculates a score for the amino acid change at a particular position. A score of <0.05 is considered as pathogenic for the protein structure. The PMUT calculates the pathological significance of non-synonymous amino acid substitution using neural networks (NN). NN output >0.5 is considered to be deleterious. PANTHER generates the substitution Position-Specific Evolutionary Conservation score. The value −3 is cutoff point for functional significance and corresponds to a Pdeleterious of 0.5. If the substitution occurs at a position not appearing in the multiple sequence alignment, a subPSEC score cannot be calculated and change is not likely to be pathogenic. The SNAP output shows prediction neutral or non-neutral, and the expected accuracy.
Previous studies have revealed a correlation between KTCN development and histopathological alterations in the structure of the corneal stroma and basement membrane, including a loss of collagen concentration [42] and rearrangement of collagen fibers [26]. Moreover, several types of collagen, including collagen type IV have been identified in the cornea [24], and COL4A1 and COL4A2 expression has been detected in the human cornea [29]. Finally, we have mapped a locus for KTCN to 13q32, in close proximity of which COL4A1 and COL4A2 are localized [21]. Given that information, we hypothesized that COL4A1 and COL4A2 are good candidates for causing KTCN in families with linkage to that locus.

Different studies have revealed several loci and a few candidate genes for familial KTCN. The first gene proposed as playing a significant role in KTCN pathogenesis was the VSX1 (visual system homeobox 1, OMIM 605020) gene. It was suggested that a few disease-causing mutations were present in this gene [43,44], but recent studies have not confirmed these findings [21,45-47]. Next, heterozygous genomic 7-bp deletion in intron 2 of SOD1 (superoxide dismutase 1; OMIM 147450) was identified in two families with KTCN [48,49]. In contrast, other studies have shown that mutations in this gene are not associated with KTCN pathogenesis [21,47]. Genetic analyses of COL4A3,COL4A4,COL8A1, and COL8A2 genes have revealed no pathogenic mutations in patients with KTCN, indicating that other genetic factors cause the disease [50-52].

We identified several single base pair substitutions in the coding regions of COL4A1 and COL4A2, including one novel heterozygous change, c.3693G>A in exon 42 of COL4A1. None of the detected alterations segregated fully with the affected phenotype in the analyzed members of the Ecuadorian KTCN families. Among the identified missense substitutions in COL4A1, one change, c.4002A>C (p. Gln1334His), was observed more frequently in KTCN patients than in healthy individuals in family KTCN-014. However, no significant statistical association of this change with familial disease could be proven (p=0.17). To predict the impact of the substitutions on the structure and function of the protein, we used different tools. All identified missense substitutions in COL4A1 and COL4A2 were predicted by the SIFT, PMUT, PANTHER, and SNAP tools to have no effect, but PolyPhen defined the Gln1334His change in COL4A1 as possibly damaging. Glutamine at this position is highly conserved in different species. Moreover, this change is present in the collagenous domain of the α1(IV) chain with Gly-X-Y repeats, which plays a role in the assembly into a triple-helical structure of the protein [22]. Replacement of the neutral residue (Gln) with the polar amino acid (His) at the Y position is likely to affect the protein structure. Nevertheless, further studies should be performed to determine the functional significance of this substitution.

To the best of our knowledge, no mutations in COL4A1 were associated with corneal disease. The spectrum of COL4A1-related disorders included porencephaly (OMIM 175780) [53-55], Hereditary Angiopathy with Nephropathy, Aneurysm and Muscle Cramps (HANAC; OMIM 611773) [56], and brain small vessel disease with hemorrhage (OMIM 607595) [57]. Recent studies have also revealed an association between mutations in exon 29 of COL4A1 and Axenfeld-Rieger anomaly with leukoencephalopathy and stroke [58]. In our study, none of the previously reported COL4A1 mutations were identified. The absence of these changes in patients with KTCN suggests that they are specific to the above-mentioned disorders only, and are not associated with KTCN in the tested families. To date, no mutations responsible for COL4A2-related human diseases have been reported.

Besides changes identified in the coding regions of COL4A1 and COL4A2, our study revealed numerous alterations in introns and UTRs of both genes, including single base pair substitutions, deletions, and insertions. Fourteen of these were novel and their clinical significance is not known. Each of the changes was observed in affected and healthy individuals in the tested families. Because important functional elements are located in non-coding regions of genes [59] and intronic alterations can result in a deleterious effect on pre-mRNA splicing [60], identification of these sequence variants could be non-accidental. Further research is needed to delineate the role of these sequence variants.

Recent studies have shown that a mouse with a mutation in a splice acceptor site of Col4a1 has ocular dysgenesis. The
mutation results in a lack of exon 40 from mice’s transcripts and leads to the accumulation of mis-folded protein in the lens epithelial cells. Col4a1Δex40 mice show optic nerve hypoplasia and anterior segment dysgenesis (ASD) including pigment
dispersion, cataracts, and corneal opacifications [61]. Splice acceptor sites are highly conserved regions in different species [56]. We detected no alterations in the splice acceptor site in intron 39 of human \textit{COL4A1}.

Extended genetic studies executed in families with KTCN have shown a high level of genetic heterogeneity [62]. The presence of many putative loci supports the hypothesis that KTCN is an oligogenic disease in which accumulation of sequence variants at several loci cause a specific KTCN haplotype and may trigger the phenotypic effect. The absence of mutations in \textit{COL4A1} and \textit{COL4A2} genes indicates that other genes are involved in KTCN pathogenesis in Ecuadorian families.

**ACKNOWLEDGMENTS**

Supported by the Polish Ministry of Science and Higher Education, Grant NN 402097837. The authors thank Genomed Company (Warsaw, Poland) for support in sequencing service.

**REFERENCES**

1. Rabinowitz YS. Keratoconus. Surv Ophthalmol 1998; 42:297-319. [PMID: 9493273]
2. Li X, Yang H, Rabinowitz YS. Longitudinal study of keratoconus progression. Exp Eye Res 2007; 85:502-7. [PMID: 17681291]
3. Cullen JF, Butler HG. Mongolism (Down’s syndrome) and keratoconus. Br J Ophthalmol 1963; 47:321-30. [PMID: 14189698]
4. Kuming BS, Joffe L. Ehlers-Danlos syndrome associated with keratoconus. A case report. Afr Med J 1977; 52:403-5. [PMID: 897848]
5. Elder MJ. Leber congenital amaurosis and its association with keratoconus and keratoglobus. J Pediatr Ophthalmol Strabismus 1994; 31:38-40. [PMID: 8195961]
6. McMonnies CW. Mechanisms of rubbing-related corneal trauma in keratoconus. Cornea 2009; 28:607-15. [PMID: 19512912]
7. Stealy LP. Keratoconus following a contact lens wear. Ann Ophthalmol 1978; 10:1177-9. [PMID: 736404]
8. Rahi A, Davies P, Ruben M, Lobascher D, Menon J. Keratoconus and coexisting atopic disease. Br J Ophthalmol 1977; 61:761-4. [PMID: 603783]
9. Edwards M, McGhee CN, Dean S. The genetics of keratoconus. Clin Experiment Ophthalmol 2001; 29:345-51. [PMID: 11778802]
10. Hughes AE, Dash DP, Jackson AJ, Frazer DG, Silvestri G. Familial keratoconus with cataract: linkage to the long arm of chromosome 15 and exclusion of candidate genes. Invest Ophthalmol Vis Sci 2003; 44:5063-6. [PMID: 14638698]
11. Tyynismaa H, Sistonen P, Tuupanen S, Tervo T, Dammert A, Latvala T, Alitalo T. A locus for autosomal dominant keratoconus: linkage to 16q22.3-2q23.1 in Finnish families. Invest Ophthalmol Vis Sci 2002; 43:3160-4. [PMID: 12356819]
12. Brancati F, Valente EM, Sarkozy A, Feher J, Castori M, Del Duca P, Mingarelli R, Pizzuti A, Dallapiccola B. A locus for autosomal dominant keratoconus maps to human chromosome 3p14-q13. J Med Genet 2004; 41:188-92. [PMID: 14985379]
13. Hutchings H, Ginisty H, Le Gallo M, Levy D, Stoesser F, Rouland JF, Arne JL, Lalaux MH, Calvas P, Roth MP, Hovnanian A, Maleczafe F. Identification of a new locus for isolated familial keratoconus at 2p24. J Med Genet 2005; 42:88-94. [PMID: 15635082]
14. Burdon KP, Coster DJ, Charlesworth JC, Mills RA, Laurie KJ, Giunta C, Hewitt AW, Latimer P, Craig JE. Apparent autosomal dominant keratoconus in a large Australian pedigree accounted for by digenic inheritance of two novel loci. Hum Genet 2008; 124:379-86. [PMID: 18795354]
15. Tang YG, Rabinowitz YS, Taylor KD, Li X, Hu M, Picornell Y, Yang H. Genomewide linkage scan in a multigeneration Caucasian pedigree identifies a novel locus for keratoconus on chromosome 5q14.3-q21.1. Genet Med 2005; 7:397-405. [PMID: 16024971]
16. Bisceglia L, De Bonis P, Pizzioloc C, Fischetti L, Laborante A, Di Perna M, Giuliani F, Delle Noci B, Buzzonetti L, Zelante L. Linkage analysis in keratoconus: replication of locus 5q21.2 and identification of other suggestive Loci. Invest Ophthalmol Vis Sci 2009; 50:1081-6. [PMID: 18978346]
17. Li X, Rabinowitz YS, Tang YG, Picornell Y, Taylor KD, Hu M, Yang H. Two-stage genome-wide linkage scan in keratoconus sib pair families. Invest Ophthalmol Vis Sci 2006; 47:3791-5. [PMID: 16936089]
18. Lisikova P, Hysi PG, Waseem N, Ebenezer ND, Bhattacharya SS, Tuft SJ. Evidence for keratoconus susceptibility locus on chromosome 14: a genomewide linkage scan using single-nucleotide polymorphism markers. Arch Ophthalmol 2010; 128:1191-5. [PMID: 20837804]
19. Hameed A, Khalig S, Ismail M, Anwar K, Ebenezer ND, Jordan T, Mehdi SQ, Payne AM, Bhattacharya SS. A novel locus for Leber congenital amaurosis (LCA4) with anterior keratoconus mapping to chromosome 17p13. Invest Ophthalmol Vis Sci 2007; 48:629-33. [PMID: 17911674]
20. Fullerton J, Paprocki F, Foote S, Mackey DA, Williamson R, Forrest S. Identity-by-descent approach to gene localisation in eight individuals affected by keratoconus from north-west Tasmania, Australia. Hum Genet 2002; 110:462-70. [PMID: 12073017]
21. Gajecka M, Radhakrishna U, Winters D, Nath SK, Rydzanicz M, Ratnamala U, Ewing K, Molinari A, Pitarka JA, Lee K, Leal SM, Bejjani BA. Localization of a gene for Keratoconus in a large Australian Caucasian pedigree identifies a novel locus for keratoconus at 2p16. Hum Genet 2002; 110:462-70. [PMID: 12073017]
corneal stroma by image analysis. Cornea 1984; 3:119-24. [PMID: 6536429]

26. Meek KM, Tuft SJ, Huang Y, Gill PS, Hayes S, Newton RH, Bron AJ. Changes in collagen orientation in distribution in keratoconus corneas. Invest Ophthalmol Vis Sci 2005; 46:1948-56. [PMID: 15914608]

27. Kenney MC, Nesburn AB, Burgeson RE, Butkowski RJ, Ljubimov AV. Abnormalities of the extracellular matrix in keratoconus corneas. Cornea 1997; 16:345-51. [PMID: 9143810]

28. Tuori AJ, Virtanen I, Aine E, Kalluri R, Miner JH, Uusitalo HM. The immunohistochemical composition of corneal basement membrane in keratoconus. Curr Eye Res 1997; 16:792-801. [PMID: 9255508]

29. Jun AS, Liu SH, Koo EH, Do DV, Stark WJ, Gottsch JD. Microarray analysis of gene expression in human donor corneas. Arch Ophthalmol 2001; 119:1629-34. [PMID: 11700013]

30. Stachs O, Bochert A, Gerber T, Koczan D, Thiessen HJ, Guthoff KM, Tuft SJ, Huang Y, Gill PS, Hayes S, Newton RH, Bron AJ. Changes in collagen orientation in distribution in keratoconus corneas. Invest Ophthalmol Vis Sci 2005; 46:1948-56. [PMID: 15914608]

31. Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. Bioinformatics 2005; 21:3445-7. [PMID: 15947021]

32. Weeks DE, Sobel E, O’Connell JR, Lange K. Computer programs for multilocus haplotyping of general pedigrees. Am J Hum Genet 1995; 56:1506-7. [PMID: 7762577]

33. Sobel E, Lange K. Descent graphs in pedigree analysis: applications to haplotype typing, location scores, and marker-sharing statistics. Am J Hum Genet 1996; 58:1323-37. [PMID: 8651310]

34. Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FC, Kennedy GC, Kong X, Murray SS, Ziegle JS, Stewart WC, Buyske S. A second-generation combined linkage physical map of the human genome. Genome Res 2007; 17:1783-6. [PMID: 17989245]

35. Thiele H, Nurnberg P, HaploPainter: a tool for drawing pedigrees with complex haplotypes. Bioinformatics 2005; 21:1730-2. [PMID: 15377505]

36. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002; 30:3894-900. [PMID: 12202775]

37. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003; 31:3812-4. [PMID: 12842254]

38. Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinformatics 2005; 21:3176-8. [PMID: 15879453]

39. Thomas PD, Campbell MJ, Kejarival A, Mi H, Karlak B, Daverman R, Diemer K, Woruganuaj A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 2003; 13:2129-41. [PMID: 12952881]

40. Thomas PD, Kejarival A, Guo N, Mi H, Campbell MJ, Munuganjian A, Lazareva-Ulitsky B. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res 2006; 34:W645-50. [PMID: 16912992]

41. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res 2007; 35:3823-35. [PMID: 17526529]

42. Sherwin T, Brooks NH. Morphological changes in keratoconus: pathology or pathogenesis. Clin Experiment Ophthalmol 2004; 32:211-7. [PMID: 15068441]

43. Héon E, Greenberg A, Kopp KK, Rootman D, Vincent AL, Billingsley G, Priston M, Dorval KM, Chow RL, McInnes RR, Heathcote G, Westall C, Sutphin JE, Semina E, Bremner R, Stone EM. VSX1: a gene for posterior polymorphous dystrophy and keratoconus. Hum Mol Genet 2002; 11:1029-36. [PMID: 11978762]

44. Bisceglia L, Ciaschetti M, De Bonis P, Campo PA, Pizzicoli C, Scala C, Grifà M, Ciavarella P, Dellé Noci N, Vaira F, Macaluso C, Zelante L. VSX1 mutational analysis in a series of Italian patients affected by keratoconus: detection of a novel mutation. Invest Ophthalmol Vis Sci 2005; 46:39-45. [PMID: 15623752]

45. Aldave AJ, Yellore VS, Salem AK, Yoo GL, Rayner SA, Yang H, Tang GY, Picornell Y, Rabinowitz YS. No VSX1 gene mutations associated with keratoconus. Invest Ophthalmol Vis Sci 2006; 47:2820-2. [PMID: 16799019]

46. Tang YG, Picornell Y, Su X, Li X, Yang H, Rabinowitz YS. Three VSX1 gene mutations, L159M, R166W, and H244R, are not associated with keratoconus. Cornea 2008; 27:189-92. [PMID: 18216574]

47. Stabuc-Silih M, Strazisar M, Hawlina M, Glavac D. Absence of pathogenic mutations in VSX1 and SOD1 genes in patients with keratoconus. Cornea 2010; 29:172-6. [PMID: 20023586]

48. Udar N, Atilano SR, Brown DJ, Holguin B, Small K, Nesburn AB, Kenney MC. SOD1: a candidate gene for keratoconus. Invest Ophthalmol Vis Sci 2006; 47:3345-51. [PMID: 16877401]

49. Udar N, Atilano SR, Small K, Nesburn AB, Kenney MC. SOD1 haplotypes in familial keratoconus. Cornea 2009; 28:902-7. [PMID: 19654524]

50. Stabuc-Silih M, Ravnik-Glavac M, Glavac D, Hawlina M, Strazisar M. Polymorphisms in COL4A3 and COL4A4 genes associated with keratoconus. Mol Vis 2009; 15:2848-60. [PMID: 20029656]

51. Aldave AJ, Bouriha N, Yellore VS, Rayner SA, Khan MA, Salem AK, Sonmez B. Keratoconus is not associated with mutations in COL8A1 and COL8A2. Cornea 2007; 26:963-5. [PMID: 17721297]

52. Stabuc-Silih M, Strazisar M, Ravnik-Glavac M, Hawlina M, Glavac D. Genetics and clinical characteristics of keratoconus. Acta Dermatovenerol Alp Panonica Adriat 2010; 19:3-10. [PMID: 20664914]

53. Breedveld G, de Coo IF, Lequin MH, Arts WF, Heutink P, Gould DB, John SW, Stone EM. Microarray analysis of gene expression in human donor corneas. Arch Ophthalmol 2001; 119:1629-34. [PMID: 9255508]

54. Gould DB, Phalan FC, Breedveld GJ, van Mil SE, Smith RS, Schimenti JC, Aguglia U, van der Knaap MS, Heutink P, John SW. Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. Science 2005; 308:1167-71. [PMID: 15905400]
55. de Vries LS, Koopman C, Groenendaal F, Van Schooneveld M, Verheijen FW, Verbeek E, Witkamp TD, van der Worp HB, Mancini G. COL4A1 mutation in two preterm siblings with antenatal onset of parenchymal hemorrhage. Ann Neurol 2009; 65:12-8. [PMID: 19194877]

56. Plaisier E, Gribouval O, Alamowitch S, Mougenot B, Prost C, Verpont MC, Marro B, Desmettre T, Cohen SY, Roullet E, Dracon M, Fardeau M, Van Agtmael T, Kerjaschki D, Antignac C, Ronco P. COL4A1 mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle cramps. N Engl J Med 2007; 357:2687-95. [PMID: 18160688]

57. Gould DB, Phalan FC, van Mil SE, Sundberg JP, Vahedi K, Massin P, Bousser MG, Heutink P, Miner JH, Tournier-Lasserve E, John SW. Role of COL4A1 in small-vessel disease and hemorrhagic stroke. N Engl J Med 2006; 354:1489-96. [PMID: 16598045]

58. Sibon I, Coupry I, Menegon P, Bouchet JP, Gorry P, Burgelin I, Calvas P, Orignac I, Doussset V, Lacombe D, Orgogozo JM, Arveiler B, Goizet C. COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. Ann Neurol 2007; 62:177-84. [PMID: 17696175]

59. Lomelin D, Jorgenson E, Risch N. Human genetic variation recognizes functional elements in noncoding sequence. Genome Res 2010; 20:311-9. [PMID: 20032171]

60. Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. J Med Genet 2005; 42:737-48. [PMID: 16199547]

61. Gould DB, Marchant JK, Savinova OV, Smith RS, John SW. Col4a1 mutation causes endoplasmic reticulum stress and genetically modifiable ocular dysgenesis. Hum Mol Genet 2007; 16:798-807. [PMID: 17317786]

62. Rabinowitz YS. The genetics of keratoconus. Ophthalmol Clin North Am 2003; 16:607-20. [PMID: 14741001]