Whole-exome sequencing in familial keratoconus: the challenges of a genetically complex disorder

Sequenciamento do exoma no ceratocone familiar: os desafios de uma doença genética complexa

Otávio de Azevedo Magalhães1,2, Thayne Woycinck Kowalski2, Gabriela Elis Wachholz2, Lavinia Schuler-Faccini2
1. Hospital Banco de Olhos de Porto Alegre, Porto Alegre, RS, Brazil.
2. Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

ABSTRACT | Purpose: The underlying genetic causes of keratoconus are essentially unknown. Here, we conducted whole-exome sequencing in 2 Brazilian families with keratoconus. Methods: Whole-exome sequencing was performed on 6 keratoconus-affected individuals of 2 unrelated pedigrees from Southern Brazil. Pathogenic variants were identified in a modified Trio analysis (1 parent and 2 children) using candidate gene filtering. All the affected subjects underwent detailed corneal tomographic evaluation. Clinically relevant variants that were present in affected individuals at minor allele frequencies <1% were examined in the 1000 Genomes Project single nucleotide polymorphism ABraOM and transcription gene (RefSeq and Ensembl) databases. Results: In family 1, a sequence variant in chromosome 1 (q21.3) was observed within the filaggrin gene. All the tested family members shared a heterozygous missense pathogenic variant in the c.4678C>T position. In family 2, exome analysis demonstrated a sequence variant in chromosome 16 (q24.2) within the gene encoding zinc finger protein 469 (ZNF469). Members of family 2 shared a heterozygous missense variant in the c.1489G>A position. In addition, the exomes of the 2 families were examined for shared genetic variants among all affected individuals. Filtering criteria did not identify any rare sequence variants in a single gene segregated in both families. Conclusion: Our findings show that a complete genotype-phenotype correlation could not be identified, suggesting that keratoconus is a genetically heterogeneous disease. In addition, we believe that whole-exome sequencing-based segregation analysis is probably not the best strategy for identifying variants in families with isolated keratoconus.

Keywords: Keratoconus; Cornea; Whole-exome sequencing; Sequence analysis; Ectasia

RESUMO | Objetivos: As causas genéticas subjacentes do ceratocone são essencialmente desconhecidas. Aqui, realizamos o sequenciamento de todo exoma de duas famílias brasileiras com ceratocone. Métodos: O sequenciamento total do exoma foi realizado em 6 indivíduos com ceratocone de duas famílias distintas do sul do Brasil. Variantes patogênicas foram identificadas em uma análise no formato de trio-modificada (um dos pais e dois filhos) usando a filtragem de genes candidatos. Todos os indivíduos afetados passaram por avaliação de tomografia de córnea. Variantes clinicamente relevantes que estavam presentes em indivíduos afetados em menores frequências alélicas <1% foram avaliadas na base de dados de polimorfismo de nucleotídeo único do 1000 Genomes Project ABraOM e do gene de transcrição (RefSeq e Ensembl). Resultados: Na família 1, uma variante de sequência no cromossomo 1 (q21.3) foi observada dentro do gene da filagrina. Todos os membros dessa família compartilhavam uma mutação missense na posição c.4678C>T. Na família 2, a análise do exoma demonstrou uma variante alélica no cromossomo 16 (q24.2) dentro do gene que codifica a proteína de dedo de zinco 469 (ZNF469). Os membros dessa família compartilhavam uma mutação missense heterozigota na posição c.1489G>A. Além disso, os exomas das duas famílias foram avaliados para variantes genéticas compartilhadas entre todos os indivíduos afetados. Os critérios de filtragem não identificaram variantes de sequência rara em um único gene segregado em ambas as famílias. Conclusão: Nossos achados indicam que uma completa correlação genótipo-fenótipo não pode ser identificada, sugerindo que o ceratocone é uma doença geneticamente heterogênea. Além disso, acreditamos que análises de segregação baseadas no sequenciamento de todo exoma provavelmente não é a melhor estratégia para identificar variantes em famílias isoladas com ceratocone.

Descritores: Ceratocone; Córnea; Sequenciamento completo do exoma; Ectasia
INTRODUCTION
Massively parallel DNA sequencing (next-generation sequencing) technology has emerged as an essential method of identifying thousands of pathogenic variants of serious conditions\(^1\). Developments in sequencing technology are now so advanced that a genome can be sequenced within a few days at a reasonable cost\(^2\). These extraordinary developments may revolutionize our ability to diagnose any disorder with a genetic background and advance understanding of the mechanisms behind many diseases\(^3\). Accurate clinical interpretation of whole-exome sequencing (WES) data, which comprises the whole protein-coding region of the genome, is complex and requires expertise in molecular diagnosis and genetic counseling in addition to bioinformatics knowledge of the patient’s suspected or diagnosed disease\(^2\).

Keratoconus (KCN) (OMIM148300) is traditionally considered a progressive non-inflammatory thinning and protrusion of the central cornea. Symptoms typically begin in puberty and may either stabilize, advance slowly to the fourth decade of life, or progress rapidly to the point of requiring a corneal transplant\(^3,4\). In recent decades, KCN has been depicted as a multifactorial disease associated with complex interactions between genetic and environmental factors contributing to manifestation of disease\(^4,5\). The identification of this disease in monozygotic twins has focused attention on its genetic component\(^6\). In spite of considering the sporadic form as the most common presentation, approximately 11% of patients have a positive family history\(^7\). A previous study reported a prevalence of KCN in 3.34% of first-degree subjects, equating with 15–65 times that of the general population\(^8\). Several surveys have shown that the most likely pattern of inheritance is autosomal dominance with incomplete penetrance or variable expressivity\(^9-11\). Numerous candidate genes have been evaluated in relation to KCN pathogenesis\(^12,13\). Nevertheless, the majority of these genes have not been confirmed as the causal agent. To explore the genetic basis of KCN, we performed WES analysis in 2 families to determine whether any genetic variants could have a significant role in KCN pathogenesis.

METHODS
This study was approved by the Federal University of Rio Grande do Sul Clinical Hospital Ethical Committee/Review Board (under protocol number 15-0559 and CAAE 49171315.0.0000.5327) and conformed to the principles of the Declaration of Helsinki. Written consent for genetic analysis was obtained from all the participants.

Clinical evaluation
Two pedigrees previously diagnosed with KCN underwent complete ophthalmic examination and corneal tomography.

Family 1 comprised a mother and her 2 daughters (Figure 1). The father was deceased, and his data were not available. Later, the mother married another man (without KCN) and had dizygotic male twins. Biomicroscopy revealed Vogt’s striae in both corneas in the oldest daughter. She had previously undergone intracorneal ring segment (ICRS) implantation in the left eye at age 23 years. The mother and the youngest daughter demonstrated a slight conical shape and corneal thinning on ophthalmic examination. Both daughters wore rigid contact lenses to correct compound myopic astigmatism. One of the twins exhibited subclinical KCN (forme fruste) on corneal tomography. The other twin had a normal corneal profile without thinning or topographic alterations.

Family 2 included a keratoconic mother and her 2 sons, both with KCN. The mother exhibited a slight conical shape and corneal thinning on ophthalmic examination. One son had previously undergone corneal transplantation in the right eye and corneal collagen crosslinking (CXL) in the left eye. The oldest son had previously
undergone CXL in the right eye, and ICRS was indicated in the left eye. The father had a normal corneal profile without thinning or topographic alterations.

Both families were considered to be of Caucasian-European descent. No relationship existed between the 2 families. Tables 1 and 2 show tomography parameters.

Exome sequencing

We performed DNA sequencing and candidate gene analyses in 3 affected members of each family in a modified Trio setting (the mothers and their clearly affected offspring). WES was performed using genomic DNA (saliva samples were extracted using an Oragene® kit) to identify the underlying genetic cause. Alignment data were evaluated using Bowtie 2 software (version 2.2.5) using the NCBI GRCh38 as a reference. Variant filtering was conducted using the pipeline Variant Annotation Analysis and Search Tool (VAAST). WES was conducted by synthesis using the Illumina HiSeq 3000 platform (UCLA Microarray Core, Los Angeles, CA, USA). Sanger sequencing of the FLG gene was performed to validate the exome and complement familial analysis using the following sequencing primers: Forward strand 5′ GTTCTGGAAGCGGACTCAG 3′ and reverse strand 5′ AGACGGAAGACACCCCTCAC 3′. Sanger sequencing of the ZNF469 gene was performed using primer pairs 5′ GTGTCAGTGACAACCTCTCC 3′ and 5′ GCGAGGTA-AGTGGGTCTTCAC 3′.

Statistical and genomic analysis

The families sequenced in the present study appeared to display a pattern of autosomal dominant inheritance; therefore, we speculated that heterozygous coding variants might explain the majority of these cases. All genotyped single nucleotide variants (SNV) were in Hardy-Weinberg equilibrium. Filtering methods were applied to establish clinically relevant variants that were present in affected individuals at global minor allele frequencies of ≤1% (GMAF ≤0.01) in the 1000 Genomes Project Single Nucleotide Polymorphism and Exome Variant Server databases. Moreover, we evaluated variants in the Brazilian ABraOM database (abraom.ib.usp.br). Briefly, a.vcf file for each individual was uploaded into the Variant Effect Predictor and Variant Annotation, Analysis, and Search Tool (VAAST 2.0)(14,15). Variant location predicted the effects on proteins across the gene expression databases (RefSeq and Ensembl), and transcription frequency or gene lengths were also noted. A p-value <0.05 was considered statistically significant.

SPSS Statistics software, version 11.0 (SPSS, Chicago, IL, USA) was used to perform statistical analyses. Searches of the OMIM and Medline databases found possible causative and relevant KCN genes. The last online search date was November 2017. The PolyPhen-2 tool was used to predict pathogenicity of a single nucleotide variant(16). The possibly damaging score for PolyPhen-2 confirmed evidence of pathogenicity. Other analyses such as SIFT and MutationTaster were included as Supplemental Digital Content (SDC).

Table 1. Clinical and tomographic features (Family 1)

<table>
<thead>
<tr>
<th>Proband</th>
<th>Ia.</th>
<th>Ib.</th>
<th>IIa.</th>
<th>IIb.</th>
<th>IIc.</th>
<th>IId.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>62</td>
<td>64</td>
<td>34</td>
<td>38</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Sex</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>BADIII ODa</td>
<td>3.44</td>
<td>0.87</td>
<td>4.28</td>
<td>7.49</td>
<td>2.14</td>
<td>1.08</td>
</tr>
<tr>
<td>BADIII ODb</td>
<td>3.29</td>
<td>0.79</td>
<td>5.04</td>
<td>*</td>
<td>2.49</td>
<td>1.33</td>
</tr>
<tr>
<td>Kmax ODa</td>
<td>48.1</td>
<td>44.5</td>
<td>47.9</td>
<td>57.6</td>
<td>45.1</td>
<td>44.1</td>
</tr>
<tr>
<td>Kmax ODb</td>
<td>47.5</td>
<td>44.3</td>
<td>48.4</td>
<td>*</td>
<td>45.4</td>
<td>44.6</td>
</tr>
<tr>
<td>Tpach ODa</td>
<td>476</td>
<td>567</td>
<td>451</td>
<td>454</td>
<td>484</td>
<td>548</td>
</tr>
<tr>
<td>Tpach ODb</td>
<td>478</td>
<td>568</td>
<td>433</td>
<td>431</td>
<td>479</td>
<td>532</td>
</tr>
</tbody>
</table>

*Data not available (submitted to Intracorneal Ring Segment).

Table 2. Clinical and tomographic features (Family 2)

<table>
<thead>
<tr>
<th>Proband</th>
<th>Ia.</th>
<th>Ib.</th>
<th>IIa.</th>
<th>IIb.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>50</td>
<td>42</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>Sex</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>BADIII ODa</td>
<td>0.51</td>
<td>4.14</td>
<td>6.27</td>
<td>*</td>
</tr>
<tr>
<td>BADIII ODb</td>
<td>0.76</td>
<td>4.24</td>
<td>11.21</td>
<td>9.25</td>
</tr>
<tr>
<td>Kmax ODa</td>
<td>43.6</td>
<td>48.2</td>
<td>48.3</td>
<td>*</td>
</tr>
<tr>
<td>Kmax ODb</td>
<td>43.5</td>
<td>48.7</td>
<td>61.6</td>
<td>54.3</td>
</tr>
<tr>
<td>Tpach ODa</td>
<td>581</td>
<td>487</td>
<td>478</td>
<td>*</td>
</tr>
<tr>
<td>Tpach ODb</td>
<td>577</td>
<td>417</td>
<td>417</td>
<td>433</td>
</tr>
</tbody>
</table>

Table 3. Single nucleotide variants identified

<table>
<thead>
<tr>
<th>Family</th>
<th>Subject</th>
<th>Number of variants</th>
<th>Depth of coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ib.</td>
<td>51,259</td>
<td>81X</td>
</tr>
<tr>
<td>1</td>
<td>IIa.</td>
<td>51,526</td>
<td>80X</td>
</tr>
<tr>
<td>1</td>
<td>IIb.</td>
<td>51,327</td>
<td>82X</td>
</tr>
<tr>
<td>2</td>
<td>Ib.</td>
<td>51,496</td>
<td>78X</td>
</tr>
<tr>
<td>2</td>
<td>IIa.</td>
<td>49,398</td>
<td>79X</td>
</tr>
<tr>
<td>2</td>
<td>IIb.</td>
<td>50,470</td>
<td>79X</td>
</tr>
</tbody>
</table>

*Data not available (submitted to corneal transplant).

Acknowledgments

This study was supported by grants from the Brazilian ABraOM database (abraom.ib.usp.br).

References

Corneal measurements

Pentacam HR Scheimpflug tomography (Oculus GmbH, Wetzlar, Germany) was used to perform corneal imaging. Tomographic parameters for each eye comprised the maximum keratometry value in diopters (D), thinnest corneal pachymetry, and Belin/Ambrosio Enhanced Ectasia display III (BAD III) “D” value. Classification of manifest KCN severity followed a BAD III value >2.6.17. Forme fruste KCN (subclinical) is considered an abortive form of KCN in which the progression process of ectasia was terminated at a certain point, most likely to recover the cornea’s biomechanical strength18. The BAD III value for this type of corneal alteration was >1.8 and <2.617.

RESULTS

We found an average of 50,913 exome sequencing variants with an average coverage of 80-fold (Table 3). Identical sequence variants (referred to as overlapping variants) were selected in all affected individuals of each family.

Molecular diagnosis in family 1 demonstrated an SNV in chromosome 1 (q21.3) within the filaggrin gene (FLG – OMIM 135940). All family members shared a heterozygous missense variant (NM_002016.1) in the c.4678C>T position (rs151103850). This SNV presented a global minor allele frequency of 0.013 in the 1000 Genomes dataset (GMAF) and an estimated frequency of 0.021 in the Brazilian population (ABraOM database), all heterozygous. PolyPhen-2 had an overall score of 0.898, predicting a possible damaging variant. Amino acid changes in the NP_001120936.2 position produced an arginine-to-cysteine substitution. Sanger sequencing confirmed variants in the FLG gene. Next, we searched for the same pathogenic variant in the twin brothers. Both boys had the same variant; in addition, one was heterozygotic (forme fruste – llc.) and the other was homozygotic (lId.). The father of the twins (lC.) did not have the FLG variant. The MutationTaster prediction score indicated that the FLG variant showed changes in amino acid, splicing, and protein sequence. Further analyses using SIFT and MutationTaster were included as SDC 3 and 4, respectively. The SIFT and MutationTaster for this variant were included as SDC 1 and 2, respectively.

In family 2, molecular diagnosis indicated a SNV in chromosome 16 (q24.2) within the gene encoding zinc finger protein 469 (ZNF469 – OMIM 612078). Family members shared a heterozygous missense variant (NM_001127464.2) in the c.1489G>A position leading to an amino acid change in the p.497G>R position (NP_001120936.2), activating a guanine to adenine substitution. This SNV (rs28723506) presented a GMAF of 0.08 and a frequency of 0.013 in the Brazilian (ABraOM database). PolyPhen-2 demonstrated an overall score of 0.912, predicting a possible damaging variant. Sanger sequencing confirmed variants in the ZNF469 gene; this rare variant was not observed in the healthy father (lA.). The MutationTaster prediction score indicated changes in amino acid, splicing, and protein sequence in this SNV. SIFT prediction of the ZNF469 gene suggests that the variant is probably tolerated. Further analyses with SIFT and MutationTaster were included as SDC 3 and 4, respectively.

We also evaluated the exomes of the 2 families for shared mutations of the genes in the exome among all affected individuals. Exome data were filtered for rare non-synonymous, coding indels, and splice acceptor and donor site variants with a GMAF ≤1% in the database. No rare sequence variants were identified in a single gene segregated in both families using these filtering criteria.

DISCUSSION

To date, this is one of the few studies to apply WES in well-characterized pedigrees affected by KCN. Heterozygous rare sequence variants were observed in 2 previously described genes, FLG and ZNF469. In this study, we identified 2 missense variants (c.4678C>T and c.1489G>A, respectively) that might promote phenotypic characterization in KCN. However, we found no common segregating rare variants among the affected members of the 2 non-related families, indicating genetic heterogeneity in the pathogenesis of KCN.

In European populations, pathogenic variants in the FLG gene have been described as a possible genetic cause of atopic dermatitis, allergic rhinitis, and asthma19. Subsequently, FLG variants were identified in Japanese, Chinese, Korean, and Taiwanese populations20. No genotype-phenotype correlation was identified in patients with FLG variants. FLG is a key epidermal differentiation protein of skin barrier function. KCN has been associated with eye rubbing and atopic diseases in various uncontrolled studies21; however, there was no correlation between serum immunoglobulin E (IgE) levels and eye complications22. Corneal thinning was...
suggested to result from mechanical trauma to the epithelium. Moreover, FLG has been identified in the central, peripheral, and limbal epithelium\[^{22}\]. Interleukin-1 (IL-1) interactions with FLG in the corneal epithelial barrier are suggested to induce keratocyte apoptosis\[^{21}\]. Multiple proteases are required for epidermal homeostasis and cleavage of IL-1 cytokines at optimal acidic pH values. A recent study found that receptor antagonist levels of IL-1α, IL-1β, IL-18, and IL-1 were increased in the uninvolved skin of patients with moderate-to-severe atopic dermatitis\[^{23}\]. This study implies a preexisting or enhanced proinflammatory status in the skin of patients with atopic dermatitis. The rs151103850 SNV has been associated with atopic dermatitis and ichthyosis\[^{24}\]. KCN and FLG have only been studied concurrently with atopic dermatitis in one report\[^{29}\]. Five patients (5.6%) with KCN and atopic dermatitis were carriers of a null mutation in the FLG gene (R501X and 2282del4). In the same article, the authors reported the need to search for other pathogenic variants of FLG. In the present study, only 1 offspring (IIa. in family 1) had a previous diagnosis of atopic dermatitis. The remaining family members had no signs or symptoms of skin disease. The interplay between FLG and KCN may result from either direct primary corneal epithelial barrier dysfunction (as well as atopic dermatitis) or indirect mechanical eye rubbing due to itching and irritation. The prediction tools agreed to indicate a potential damaging variant. It is noteworthy to emphasize that FLG protein formation is not invalidated by the existence of homozygosity. Sanger sequencing of the twin brothers in family 1 revealed that they carry the same FLG variant; however, 1 twin was homozygous (Ilc.) and the other was heterozygous (Ild.). Conversely, the mild phenotype (forme fruste) was only expressed in 1 subject (IIc.). That this variant was identified in a healthy brother along with the absence of skin conditions in the family may indicate that variations in the FLG gene might not be responsible for KCN unless incomplete penetrance occurs.

ZNF469 (OMIM 612078) is a single exon gene encoding a 413-kDa protein composed of 3953 amino acid residues containing 5 different C2H2 zinc finger domains in its C-terminal, suggesting its role in the transcription process\[^{22}\]. Protein homology suggests that ZNF469 could function as an extra-nuclear regulator factor for synthesizing and organizing collagen fibers, constituting the major component of the human cornea\[^{26}\]. This gene has 30% sequence homology to the helical regions of COL1A1, COL1A2, and COL4A1, all of which are highly expressed in the cornea\[^{27}\]. ZNF469 was identified as the gene responsible for Brittle cornea syndrome type 1 (BCS1, OMIM 229200), an autosomal recessive disorder characterized by an extremely thin and fragile cornea, prone to spontaneous rupture\[^{26,27}\]. Recent genome-wide association studies have shown that variations in ZNF469 could also contribute to central corneal thickness (CCT). Based on an evaluation of 5 cohorts from Australia and the United Kingdom, the authors reported an association between CCT and 2 SNPs (rs12447690 and rs9938149) mapped in the intergenic region upstream to ZNF469\[^{28}\]. Other authors have described significant enrichment of potentially pathologic heterozygous ZNF469 alleles in 12.5% of KCN individuals of European ethnicity, most likely making it the most significant genetic factor in KCN\[^{26}\]. A subsequent study performed in a Polynesian population also reported a rare missense variant in ZNF469 in 23% of KCN patients\[^{26}\]. Not all studies agree that ZNF469 variants segregate in families with KCN. Other authors have reported that the presence of heterozygous loss-of-function alleles in ZNF469 did not influence the development of KCN\[^{29}\]. Therefore, prediction methods disagree, and it is too early to define whether variants in ZNF469 are causative for KCN. Our modified Trio analysis suggests a potential role for ZNF469 in Brazilian individuals with KCN.

WES has emerged as a powerful tool for systematically exploring rare coding variations\[^{30}\]. Since the most known genetic causes of diseases affect protein-coding regions, the exome is a logical place to look for potentially causative variants in disorders exhibiting Mendelian inheritance. We retained only variants within exonic regions for this experiment. Non-coding variants, epigenetic changes, and epistatic interactions might be important in KCN development and in other complex diseases, in which case alternative study designs should be utilized. Additionally, most variants identified by WES are not clinically significant, and our lack of understanding of genome biology will continue to compromise our ability to interpret its results. Analytical validity is limited because bioinformatics algorithms for interpreting real data are still underway.

In the past, subclinical forms of KCN were a confounding factor, complicating the identification of a correct phenotype\[^{9}\]. Recent advances in corneal tomography for detecting KCN, including forme fruste, demonstrate higher accuracy in delineating study populations\[^{7}\]. The use of subclinical phenotypes can have a significant effect in genetic studies in the presence of reduced
Whole-exome sequencing in familial keratoconus: the challenges of a genetically complex disorder

penetrance or variability in phenotype expression, and the use of subclinical phenotypes might allow us to identify the abnormal genotype in the absence of clinical disease. As with other diseases of complex etiology, differentiating between association, cause, and effect is challenging and varies between individuals.

Therefore, we believe that WES-based segregation analysis is probably not the best strategy for identifying pathogenic variants in isolated KCN families. Current understanding supports that KCN is caused by multiple genes and, in many instances, may result from complex interactions between genes and environmental factors, such as eye rubbing and atopy of the eye\(^\text{100}\). Affected individuals might thus have KCN due to the presence of rare sequence variants of risk factors and not merely to direct anatomical defects. The best chance to identify genes is in rare, large single families or in rare populations with high concentrations of KCN due to a common founder.

Even though probable damaging variants were identified in the \textit{ZNF469} and \textit{FLG} genes, further analysis of these candidate genes in larger cohorts is required to confirm their involvement in the development of KCN and in the phenotypic variability between the analyzed individuals. In addition, no evidence of rare sequence variants was found in a single gene that segregates with KCN in these 2 non-related Brazilian families, suggesting that the disease is genetically heterogeneous.

In conclusion, this study reports on the difficulty in evaluating rare variants in complex disorders such as KCN. We present evidence to confirm genetic heterogeneity in KCN pathogenesis rather than a single major gene effect. Further screening in controls without KCN for both variants in the same population and replication with larger multiethnic familial and sporadic samples remains necessary.

**REFERENCE**


