International Journal of Frontiers in
Life Science Research
Journal homepage: https://frontiersrj.com/journals/ijflsr/
ISSN: 2783-0470 (Online)

(RESEARCH ARTICLE)

Check for updates

Transcriptome RNA-Seq Analysis of Normal and Keratoconus Corneal Epithelium

Fatma Hassan Boshahma [1, 2], Muna Salem M. Solyman [3], Manir Ali [4], Euan Andrew McDonnel [2], Monder Wail Elhawari [5] and Abeer Hussein Amer [1, 6, *]

[1] *Department of Clinical Laboratory Sciences, Faculty of Applied Medical Science, Libyan International Medical University, Benghazi, Libya.*
[2] *Department of Biological Sciences, Faculty of Biological Sciences, University of Leeds, Leeds, UK.*
[3] *Department of Molecular Diagnostic, Faculty of Biomedical Sciences, University of Benghazi, Libya.*
[4] *Department of Medicine, School of Medicine, University of Leeds, Leeds, UK.*
[5] *Department of Medicine, Faculty of Medicine, Libyan International Medical University, Benghazi, Libya.*
[6] *Department of Histology, Faculty of Medicine, University of Benghazi, Benghazi, Libya.*

International Journal of Frontiers in Life Science Research, 2024, 06(02), 027–043

Publication history: Received on 25 March 2024; revised on 08 May 2024; accepted on 11 May 2024

Article DOI: https://doi.org/10.53294/ijflsr.2024.6.2.0032

Abstract

Background: Keratoconus (KTCN) is a progressive eye condition characterised by thin, misshapen corneas arising in the teens or early twenties, and can cause visual disability. While this can be corrected with glasses or contact lenses, collagen cross-linking of the cornea slows disease progression. In advanced disease, patients undergo transplantation of the corneal epithelial layer with normal donor tissue. The cause of KTCN is unknown, but there is evidence for both environmental and genetic bases.

Material and Methods: RNA-Seq data from published work by Kabza et al., 2017, was used to investigate the pathology of cornea tissue, analysing twenty-five KTCN and twenty-five non-KTCN samples from RNA-seq experiments. Principal component analysis plots were used to identify how characteristics impact development of keratoconus.

Results: The findings show that KTCN is more common in teenagers and young adults, and all 25 KTCN samples were for individuals under 30 years old. Principal component analysis also showed that approximately 35% of samples were from individuals who rubbed their eyes. It also showed that KTCN is more common in males, representing over 65% of samples. Differential expression analysis of RNA-sequencing data was used to identify differentially expressed genes, downregulated genes were found to be enriched in many pathways, including cytokinecytokine receptor interaction pathways, and played a role in keratoconus development.

Conclusion: Upregulated genes were enhanced in six pathways, including the peroxisome pathway. A potential relation was found between defects in peroxisome and keratoconus development. Overexpression of peroxisome genes increases phytanic acid levels, and causes Zellweger spectrum, affecting the cornea and potentially leading to keratoconus.

Keywords: Keratoconus; Differential expression analysis; Pathway overrepresentation analysis; Upregulated and downregulated genes; Peroxisome genes

* Corresponding author: Abeer H. Amer

1. Introduction

Keratoconus (KTCN, OMIM 148300) is bilateral eye disorder characterised by progressive thinning and changing of the shape of the cornea (Krachmer et al, 1984; Khaled et al, 2017). Progressive thinning of the cornea results in developing a cone-shaped bulge toward the centre in the areas of greatest thinning of the cornea (Khaled et al., 2017). This cone shape of the cornea causes problems in focusing light appropriately on the retina. This leads to the development of blurry vision, sensitivity to light and loss of vision in some cases (Krachmer et al., 1984). Despite KTCN having been identified as a non-inflammatory condition, recently, it has been identified that KTCN can be an inflammatory condition. One such study showed that the tears of people with KTCN had a high level of cytokines compared with tears from controls. The protein expression level of IL-4.-5,-6 and TNF-α,-β, with a function in regulating ocular surface inflammation, was very high in tears from KTCN patients (Balasubramanian et al., 2012, Song et al., 2024).

The age at which keratoconus emerges varies between late teenage years and young adulthood (Kymes et al., 2004, Godefrooij et al., 2017). However, other studies give the age of diagnosis, in which the mean diagnostic age was from 20 years to 24.05, as reviewed by (Gordon-Shaag et al., 2015). The causes of keratoconus are still not fully clear. While most KTCN appears sporadic, the condition can be caused by family history, in which cases of autosomal dominant and autosomal recessive modes of inheritance have been determined (Wang et al., 2000; Tyynismaa et al., 2002; Hughes et al., 2003). It has been shown that there is an association between certain systemic diseases and increased risk of KTCN. For example, persons with Down's syndrome, sleep apnoea, and asthma have an increased risk of keratoconus and require specific care (Woodward et al., 2016).

The development of KTCN in some individuals might be related to environmental influences such as eye rubbing, contact lens wear or atopy of the eye (Abu-Amero et al., 2014). In addition, extensive exposure to strong sunlight or ultraviolet light leads to oxidative damage. KTCN could also be related to genetic inheritance (Abu-Amero et al., 2014). Also studies showed that keratoconus severity is more concordant in monozygotic MZ than in dizygotic DZ, and that the development of keratoconus is contributed to by genetic components (Tuft et al., 2012). Other studies done in families also show that genetic effects contribute to keratoconus (Wang et al., 2000, Lin et al., 2023).

KTCN distribution is not equal in all countries worldwide. The prevalence of KTCN is 1.2% in the Australian population and 0.6% in Japan population which is very low (Hofstetter, 1959, Chan et al., 2021). However, KTCN prevalence is 2.3% in Maori population, 3.3% in Iranian population and 2.3% in Indian population, because people are exposed more to sunlight, leading to oxidative damage (Hashemi et al., 2013, Papali'i-Curtin et al., 2019, Chan et al., 2021).

Linkage analysis has determined the chromosomal regions of this disease. Multiple genetic loci have been identified using linkage analysis in families. The first identified locus was on chromosome 20p11.21 and it is caused by a heterozygous mutation in the VSX1 gene (MIM605020) (Héon et al., 2002). The second causative gene is located on chromosome 16qq22.3– q23.1 between the markers D16S2624 and D16S3090 (Tyynismaa et al., 2002). The third locus causative gene was identified to chromosome 3p14-q13 (Brancati et al., 2004). The fourth region was identified to cause KTCN is on chromosome 2p24 (Hutchings et al., 2005). The fifth locus was identified to be located on chromosome 5q14.1-q21.3 (Tang et al., 2005). The sixth locus was identified to be on chromosome 9q34 (Li et al., 2006). The seventh locus that cause KTCN was found to be on chromosome 13q32 (Gajecka et al., 2009). The eighth locus was identified in 2010 and is located on chromosome 14q24.3 (Liskova et al., 2010, Kondkar et al., 2023). The final locus is located on chromosome 2q21.1 and is caused by a TUBA3D gene (MIM617878) (Hao et al., 2017). Association Study Recently, a genome-wide association study (GWAS) of keratoconus was performed on 6,669 cases and 116,547 controls. It revealed 31 genomic loci associated with impairment of corneal collagen matrix integrity and pathways for differentiated cells (Hardcastle et al., 2021). These include: ATP1B1, MRPS14, CD34/CD46, LRP1B, FNDC3B, ITGA2, LOX, TFAP2B, COL12A1, NDUFAF6, ACTL7B, COL5A1, FBXW5, EIF3A, PIDD, FAM76B, GRIN2B, GALNT6, FOXO1, KLF5, TNFAIP3L3, RORA, ZNF469, KIF1C, ALDH3A1, RAB11FIP, SKAP1, COL1A1, STK35, COL6A1, and AIFM3 (Hardcastle et al., 2021).

Aim

The aim of the research was to analyse RNA-seq datasets to identify the pathway analysis and determine differential expressed genes and their role in developing Keratoconus.

2. Materials and Methods

This research was performed based on a previous study by (Kabza et al. (2017). The data contains RNA-Seq samples, 25 with disease and 25 healthy, as Gene Expression Omnibus (GEO) read count matrices. Basically, these are in a form where they have already been aligned and quantified, and data are tabulated with samples shown as columns as samples

and genes as rows. The data were obtained from a previous RNA-Seq study performed for KTCN samples obtained from Polish patients during the keratoplasty procedure, and for non-KTCN samples used as a control, from patients who had been referred for corneal transplantation for different reasons (GSE77938) (Kabza et al., 2017). Before participation in the study, each individual provided written informed consent in line with the Declaration of Helsinki. The protocol for the study was approved by the Institutional Review Board at Poznan University of Medical Sciences, Poznan, Poland. As described in Kabza et al. (2017), all RNA samples from all tissues were extracted and purified with a Total RNA Purification Kit (Norgen Biotek) (Kabza et al., 2017).

To access the data, the GEOquery Bioconductor R package was used to download the data into RStudio version 4.0.3 on a PC, using the access code GSE77938 from the paper (Davis and Meltzer, 2007; Kabza et al., 2017). The function data.frame was used to transfer the metadata of phenotype data GSE77938 into RStudio. Two files (GSE77938_discovery_gene_counts.txt.gz) and (GSE77938_replication_gene_ count.txt.gz) were downloaded to PC from the supplementary files on the GEO website (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77938), using the GEO accession code (GSE77938), and then uploaded to RStudio. The function data.frame was also used to combine these two files together. Finally, the row names of the metadata table were matched with the column names of matrix data to keep them in the same order. The org.Hs.eg.db package was used to map the gene symbol into its Entrez identifier (Carlson, 2019). The Columns method in the package was used to extract the "SYMBOL" (gene symbol), "GENE NAME" information and "ENTREZID" for each gene. The Keytype method in the same package was used to extract ENSEMBL identifiers. Additionally, the Select method was used to specify what columns and keys we want to extract.

Differential expression is important for analysis of RNA sequencing data, and functions by finding the differentially expressed genes among a group of samples (Love et al., 2014). It helps to determine which genes were dysregulated in KTCN in this research. The DESeq2 package was used to store the read counts (Love et al., 2014). The function DESeqDataSetFromMatrix was used with the count data feature for the metadata, and the colData feature for the matrix data, to provide information about the samples. It also includes the design formula for indicating both the KTCN and non-KTCN conditions and replication and discovery experiments, to construct DESeqDataSet object data and convert counts to integer mode. Then the function DESeq was run on the constructed DESeqDataSet object for estimation of the size factor and the dispersion of counts for each gene. Finally, the function variance stabilising transformations (vst), was used to extract the normalised matrix data from the DESeqDataSet object.

The principal component analysis (PCA) tools package was used with normalised matrix data to form plots (Blighe and Lun, 2021). The function plot PCA from this package was used with the ingroup feature to determine which columns should be examined. The function ggplot from the ggplot2 package was used to plot the PCA plots (Wickham,2016). Firstly, a PCA plot was constructed between cases and controls. Secondly, three separate PCA plots were plotted between the two groups, with the features: eye rubbing, age and sex.

To analyse and generate hierarchical clustering of KTCN and non-KTCN samples with differentially expressed genes, the Pretty Heatmaps (Pheatmap) package was used to run a heatmap between two samples (Kolde, 2019). The normalised matrix data was used to form the heatmap plot. The top 100 and 1000 differentially expressed genes were chosen for the heatmap plot, because this method does not work for thousands of genes. The annotations package was used with the heatmap package to place the name of the genes in the rows, while columns represented samples: either cases or controls; and showed hierarchical clustering on samples and genes.

Change shrinkage Log Fold Change (LFC) was calculated for the DESeqDataSet object data using the function lfcShrink with respect to KTCN vs non-KTCN samples and apeglm type, to shrink the LFC in order to visualise and rank genes. Then, the most significant genes with an adjusted p-value of less than 0.01 were chosen from these data to be used in the next steps in the process. The annotations package was used in this step to obtain the symbol of genes from ensemble ID. The function summary was run to obtain the percentage of genes that are upregulated and downregulated. Finally, the MA-plot was plotted to explore the results that got after calculating the LFC. The function plotMA was run to show the log fold changes against the mean of the normalized counts for all samples. The EnhancedVolcano package was used to plot the volcano (Blighe and Lun, 2021). The results from the differential expression genes were used in this step.

Genes with an adjusted p-value of less than 0.01 were used to find the pathway analysis. These data were split into downregulated genes with a log fold change less than zero and upregulated genes with a log fold change higher than zero. The annotations package was used in this step to obtain the symbol of genes from ensemble ID. These symbols of genes for upregulated and downregulated data were then submitted separately in the submission box for the Enricher website https://maayanlab.cloud/Enrichr/. After submission, the pathway tool was used to check which pathways these up- or downregulated genes are involved in. Also, the Ontologies tool on the Enrichr server was used to identify

which processes their proteins were related to. The Disease/Drug tool was used to check what diseases could be caused by these deregulated genes.

3. Results

Principal component analysis (PCA) has a role in detecting differentially expressed genes. A PCA plot was generated for twenty-five non-KTCN and twenty-five KTCN samples, to check the variation between the two groups and identify whether they had been successfully clustered. It resulted that five samples from non-KTCN samples (KR-53, KR_57, KR_64, KR_08 and KR_35) clustered with KTCN samples rather than with non-KTCN samples (Figure 1a). After removing these five samples resulted in a good separation between KTCN and non KTCN groups as it is clear in the PCA plot (Figure 1b). These samples were removed also from further analysis.

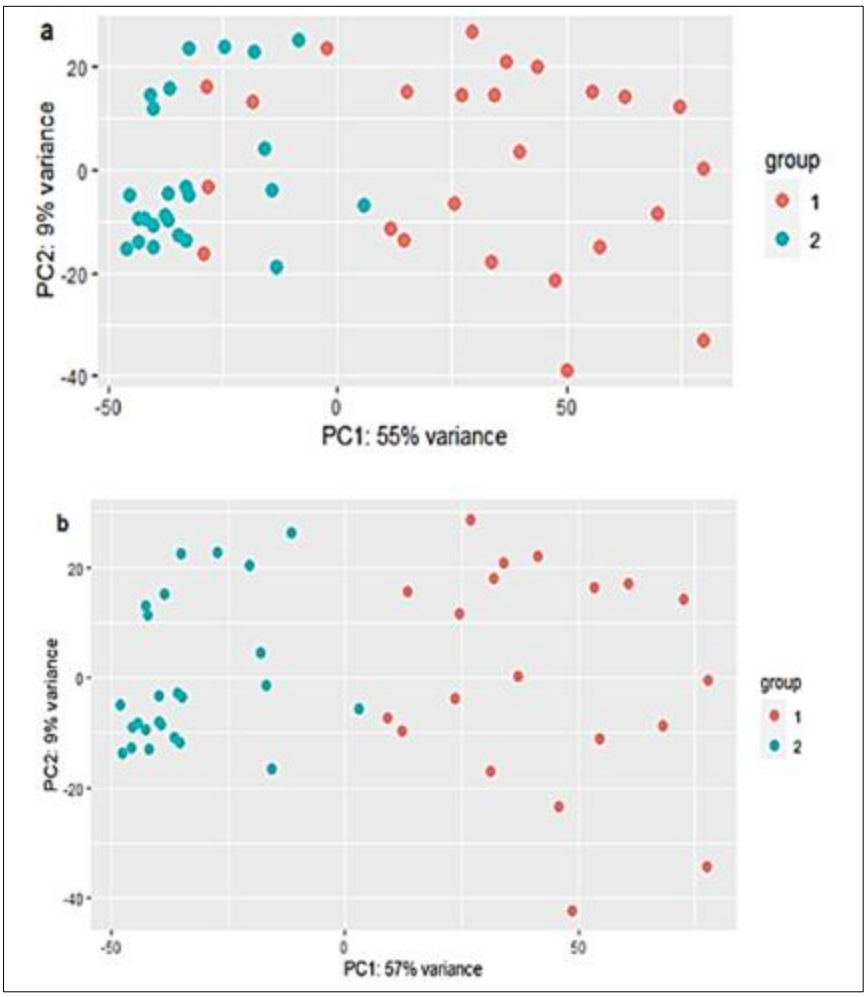


Figure 1 PCA plots of KTCN and non-KTCN samples. Two plots show that the variation is from left to right. Individuals that have similar transcripts are clustered together. PCA plots based on the expression of genes before removing the samples with PC1 55% and PC2 9% variance (a) and after removing the samples with PC1 57% and PC2 9% variance (b). PC1 (First Principal Component) is the axis that spans the most variation. PC2 (Second Principal Component) is the axis that spans the second most variation. Points colour labelling shows non-KTCN samples in pink (1) and KTCN samples in turquoise (2)

A PCA plot was used to plot the KTCN and non- KTCN samples with the age of the individual, to check the correlation between the disease and age. This research used the age at which the sample was taken from the individuals. The PCA plot showed that age for most of the twenty-five KTCN samples was less than thirty (Figure 2). This means that KTCN could be developed in younger adults and that there could be a correlation between disease state and age.

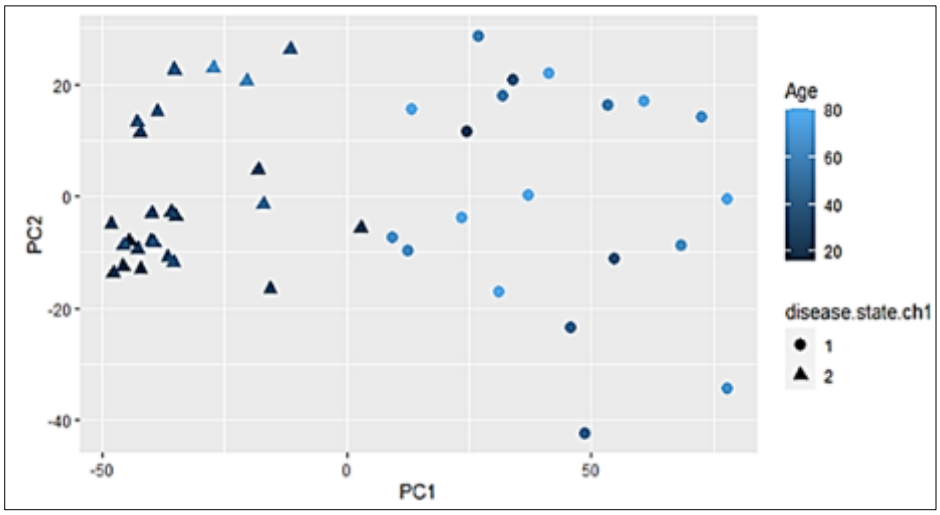


Figure 2 Principal component analysis plot. The plot shows that the variation is from left to right between KTCN and non-KTCN samples with age feature. The Age from 20 years old to 80 years old is ranged from dark blue to light blue. The triangle symbol (▲) represents KTCN samples and circle symbol (●) represents non-KTCN samples

The correlation between KTCN and eye rubbing was also checked. When plotting the eye rubbing factor with KTCN and non-KTCN samples, the results showed that there are nine individuals of KTCN group rubbing their eyes and only one individual of non-KTCN group rubbing his/her eye (Figure 3). It means that individuals who are rubbing their eyes, could have a higher chance for developing KTCN. It showed that might be a correlation between the disease state and eye rubbing.

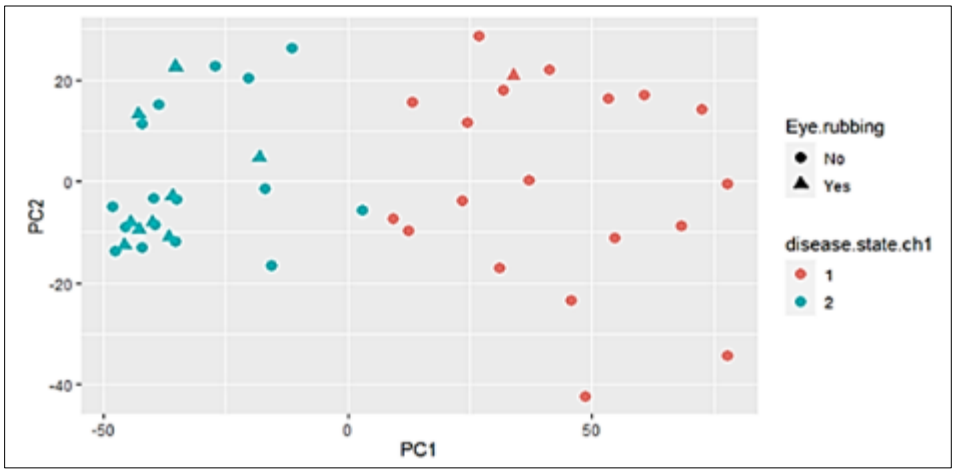


Figure 3 Principal component analysis plot. The plot shows variation from left to right between KTCN and non-KTCN samples with and without eye rubbing. The circles (●) represent individuals who do not rub their eyes (No), and the triangles (▲) represent individuals who rub their eyes (Yes). Point colour labelling shows non-KTCN samples in pink (1) and KTCN samples in turquoise (2)

To identify whether sex factor has a role in developing KTCN, a PCA plot was generated for sex across KTCN and non-KTCN samples. The results showed that there are seventeen male individuals from the twenty-five KTCN samples (Figure 4). This suggests that males might be more commonly affected by KTCN than females. This is could be a correlation between the disease state and sex.

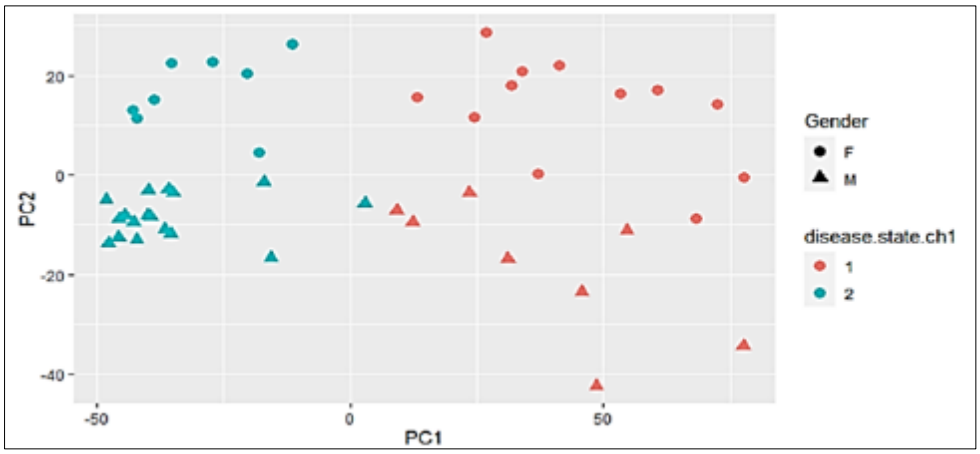
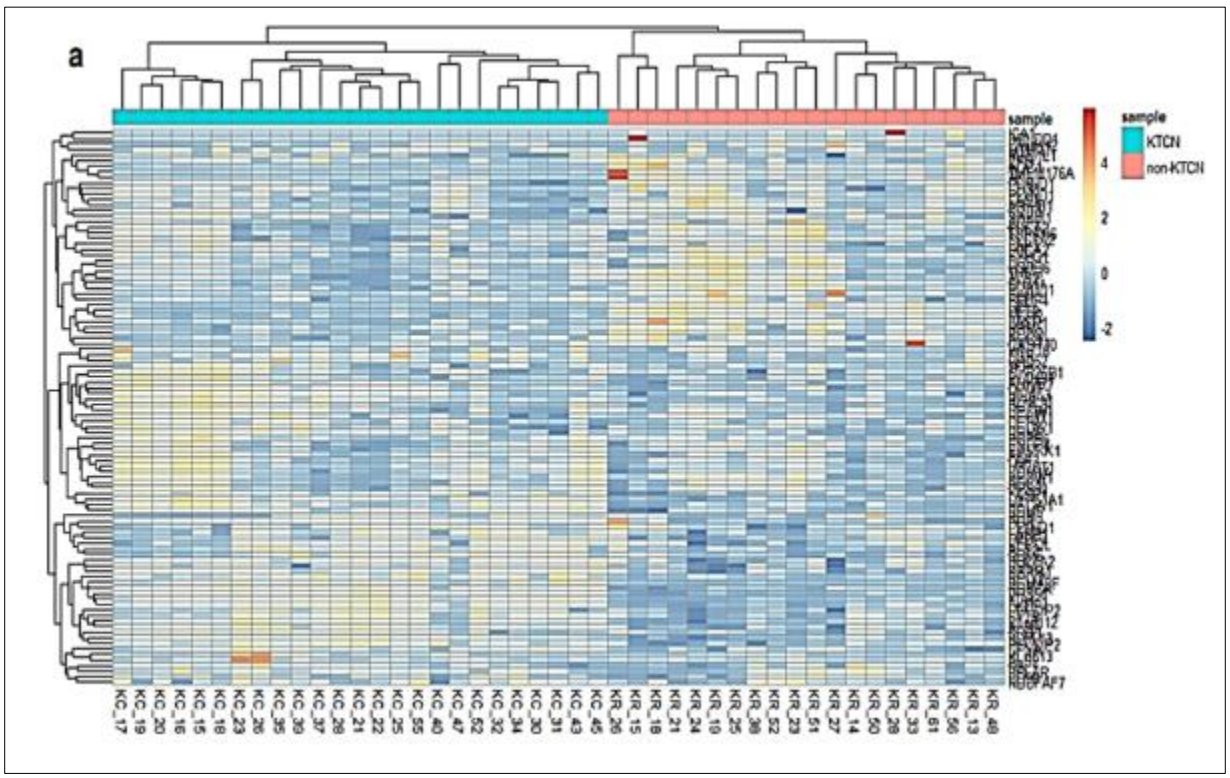


Figure 4 Principal component analysis plot. The plot shows variation from left to right between KTCN and non-KTCN samples by sex. The circle (●) represents individuals who are female (F) and the triangle (▲) represents individuals who are male (M). Points colour labelling shows non-KTCN samples in pink (1) and KTCN samples in turquoise (2)

Hierarchical clustering analysis is based on representing the data between samples as a colour image. The rows and columns are reordered so that samples with similar transcripts/profiles are much closer to one another, making these transcripts more visible. Each value in the data matrix is represented as a colour, which makes it easy to view patterns in graphic form. The clustering which occurred based on the top 100 and 1000 differentially expressed genes showed a clear separation between KTCN and non-KTCN individuals (Figure 5a, b).



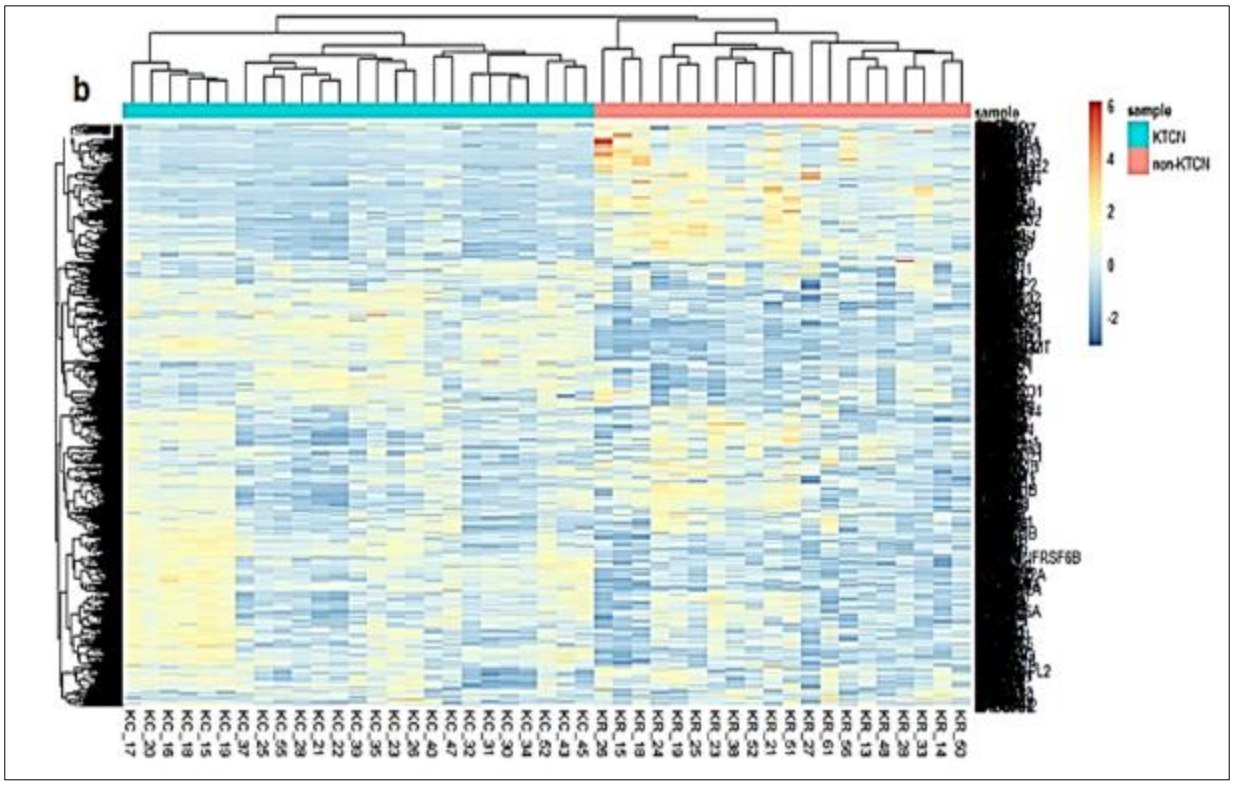


Figure 5 Hierarchical clustering analysis of expressed genes. A heat map shows a hierarchical clustering of KTCN and non-KTCN samples based on expression of the top 100 genes (a) and top 1000 genes (b). Data is separated into groups of genes on the vertical line and groups of individuals on the horizontal line. The colour scale on the right presents in which blue means that genes are downregulated, and orange/red means that genes are upregulated. Upper colour labelling shows KTCN samples in turquoise and non-KTCN samples in pink

After calculating LFC shrinking and extracting those most significant genes with an adjusted p-value of less than 0.01, the summary results show that approximately 13,479 genes are significantly differentially expressed. The number of genes that are significantly downregulated was found to be higher than for genes that are significantly upregulated in KTCN individuals, at 54% and 46% respectively (Table 1).

Table 1 The total number of genes that are differentially expressed. LFC is Log Fold Change. Genes that have a LFC value of more than zero are upregulated and genes with a LFC value of less than zero are downregulated. The number of genes that are significantly differentially expressed at adjusted p-value 0.01 is 13,479. 46% of those genes were upregulated, while 54% were downregulated in the KTCN samples

Type of Genes	**Total Read Count (adjusted p-value <0.01)**	**Percentage**
LFC >0 (up-regulated genes)	6178	46%
LFC <0 (down-regulated genes)	7301	54%

Extracted genes with an adjusted p-value of less than 0.01 were used to produce an MA-plot. This plot is often used in analysis of gene expression data from microarray and RNA-Seq experiments to measure change in the expression level of a gene. The results showed that all of genes are more significant and there are no low significant genes. The plot also showed the number of significantly upregulated genes is lower than the number of significantly downregulated genes (Figure 6). The MA-plot confirmed results that had been obtained from Table 1.

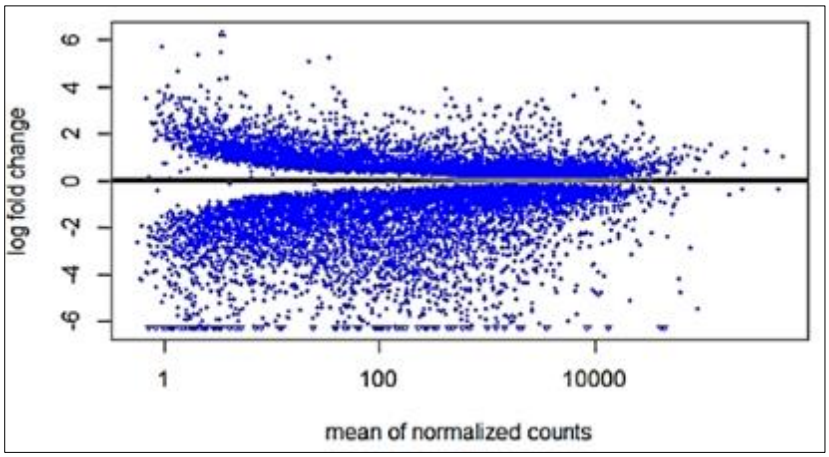


Figure 6 The MA-plot. Genes above the grey/ black line (the line at zero for the log fold change on the y-axis) are upregulated, and below the grey/ black line are downregulated. Blue plots mean that genes are more significant. Genes that have an LFC value of more than zero are upregulated, and genes with an LFC value of less than zero are downregulated

An Enhanced volcano plot is a type of scatter plot that shows the p-value, which gives statistical significance, against the fold change, which is the magnitude of change. The enhanced volcano plot identifies genes that have a large fold change and are statistically significant (Figure 7). The plot also showed that there are more genes that are downregulated than upregulated, and that there are more genes downregulated which are statistically significant (Figure 7).

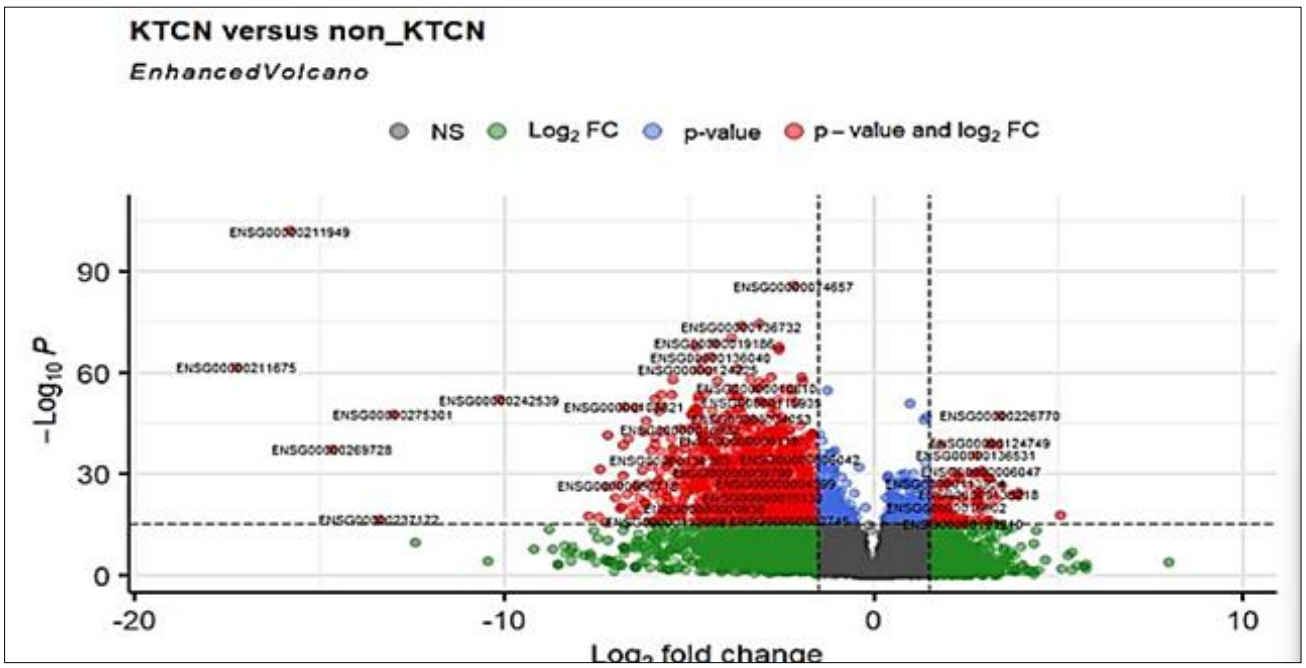


Figure 7 Enhanced volcano plot. The ensemble ID of genes that are coloured. The coloured genes are those which have passed the thresholds for p-value or Log2 Fold Change or both (green, blue and red), while in the grey region, genes have not passed the threshold. Green indicates absolute Log2 FC > 1 but p < 0.01. Blue is p < 0.01 but absolute Log2 FC < 1. Red is p < 0.01 AND absolute Log2 FC > 1. The most upregulated genes are towards the right, while the most downregulated genes are towards to the left

Analysis of molecular pathways was performed on the separate sets of upregulated and downregulated genes. Downregulated genes based on splitting genes with a log fold change, the number of downregulated genes is 5,908. These downregulated genes showed that they were enhanced in 95 different pathways at p-value < 0.01. The top 10 pathways showing the most significant KTCN overrepresentation involve cytokine-cytokine receptor interaction, coronavirus disease, and viral protein interaction with cytokine and cytokine receptor and other 7 pathways (Figure 8).

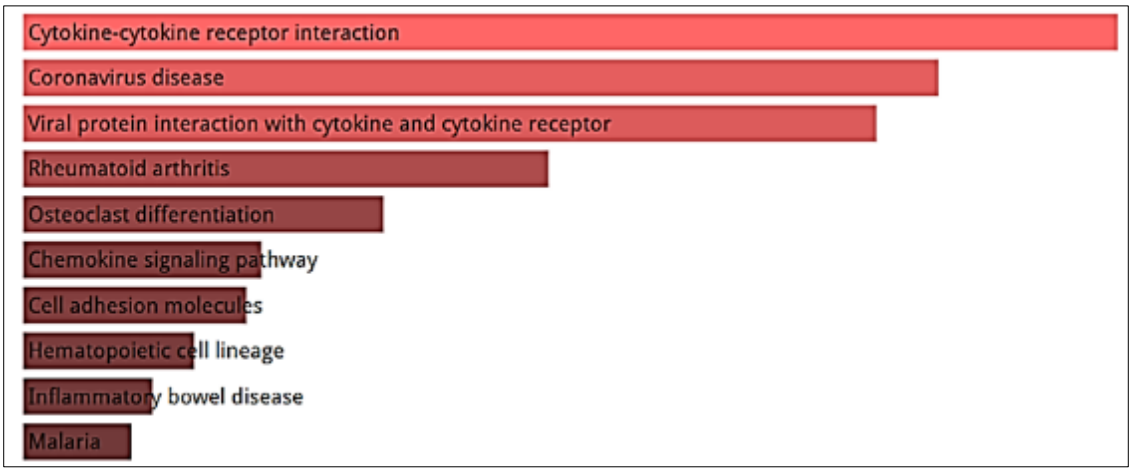


Figure 8 The top KEGG 2021 pathways overrepresented across differentially expressed genes for downregulated genes, detected by the Enrichr server (p-value <0.01). The cytokine-cytokine receptor interaction pathway is the pathway most related to the downregulated genes for KTCN. The lighter the colour, the lower the p-value of the pathway

The clustergram tool from the Enrichr server was used to obtain the top 40 downregulated genes which were enhanced in 10 pathways, including IL6, TNF, IL1B and many other genes for cytokine-cytokine receptor interaction and other top nine pathways (Figure 9).

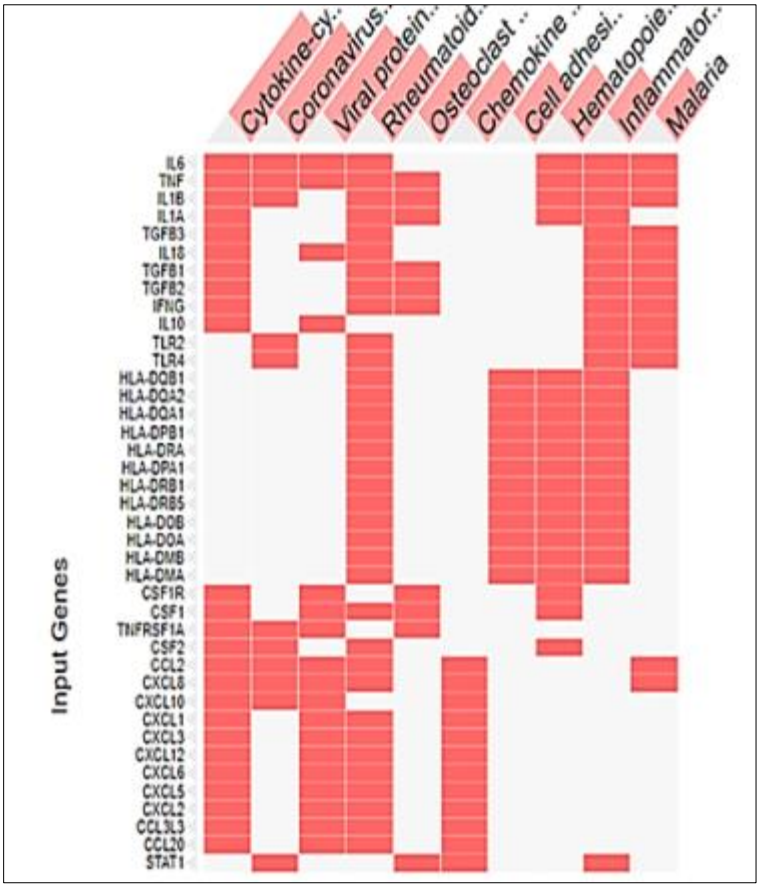


Figure 9 The clustergram diagram showing downregulated genes involved in 10 pathways in KTCN corneas. It detected by Enrichr server (p-value <0.01). The rows represent genes. The columns represent pathways. Pink boxes show genes that has a role in the pathways

To identify the properties of the gene products for all 5,908 downregulated genes, the Ontology tool in the Enrichr server was used to check the Gene Ontology cellular component, GO molecular function and GO biological process. The results showed that GO is different from one domain to another in the downregulated genes (Figure 10).

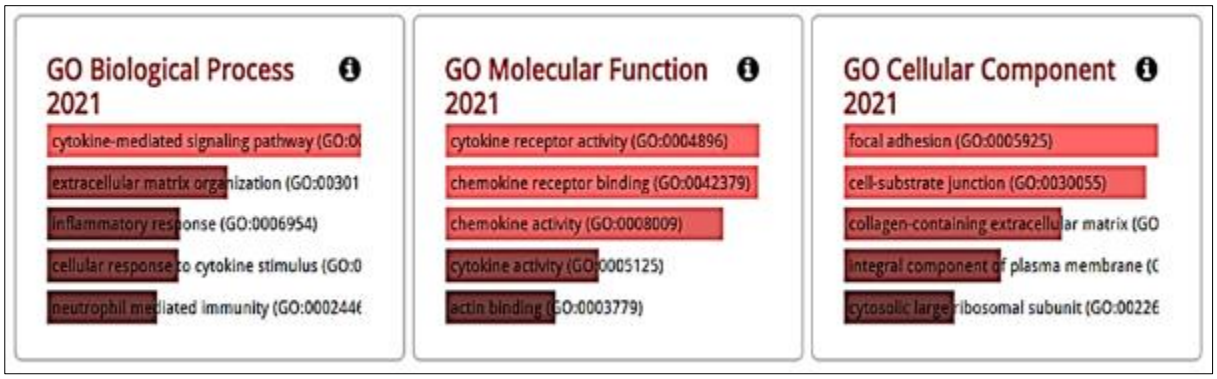


Figure 10 The Gene Ontology. It shows Ontology domains: GO biological process (a), GO molecular function (b) and GO cellular component (c), detected by the Enrichr server (p-value <0.01). The cytokine-mediated signalling pathway has the most GO biological processes in downregulated genes for KTCN. Cytokine receptor activity is the most prominent GO molecular function in the downregulated genes for KTCN, while focal adhesion is the most frequent GO cellular component. The lighter the colour, the lower the p-value for the function

The Diseases/Drugs tool in the Enrichr website was also used to check which diseases were related to the identified 5,908 downregulated genes. There are thousands of diseases that are related to these genes at a p-value <0.01, including liver cirrhosis and rheumatoid arthritis, which are the diseases most related to the genes (Figure 11).

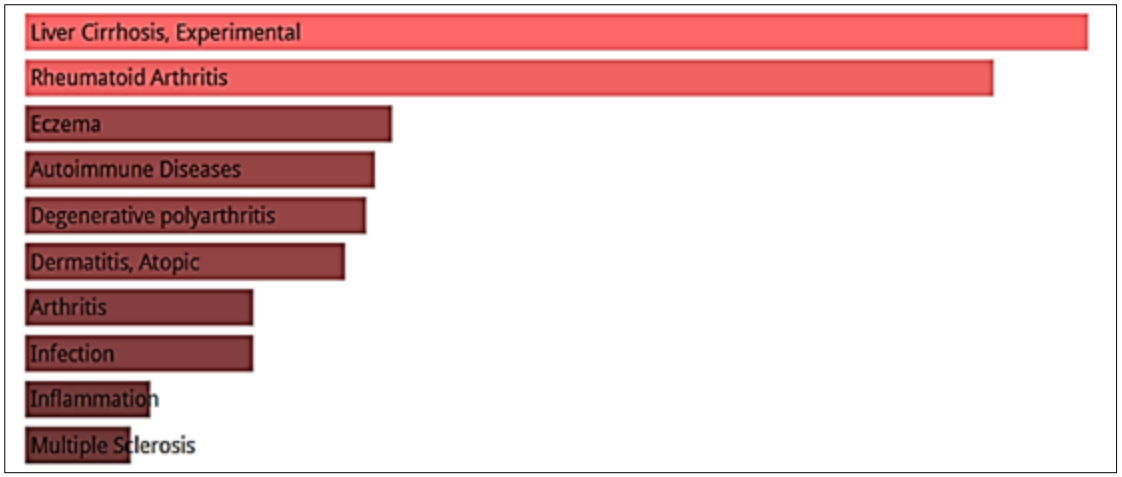


Figure 11 Diseases caused by the downregulated genes, detected by Enrichr server (p-value <0.01). The diseases in the graph are the most common diseases for the downregulated genes. Experimental Liver Cirrhosis has the lowest p-value, followed by Rheumatoid Arthritis. The lighter the colour, the lower the p-value for the disease

Based on splitting genes with a log fold change, the number of upregulated genes is 4,444. These upregulated genes showed that six pathways are more significant in KTCN in terms of overrepresentation, involving Herpes simplex virus 1 infection, valine, leucine and isoleucine degradation, propanoate metabolism Selenocompound metabolism, peroxisome and pyruvate metabolism (Figure 12).

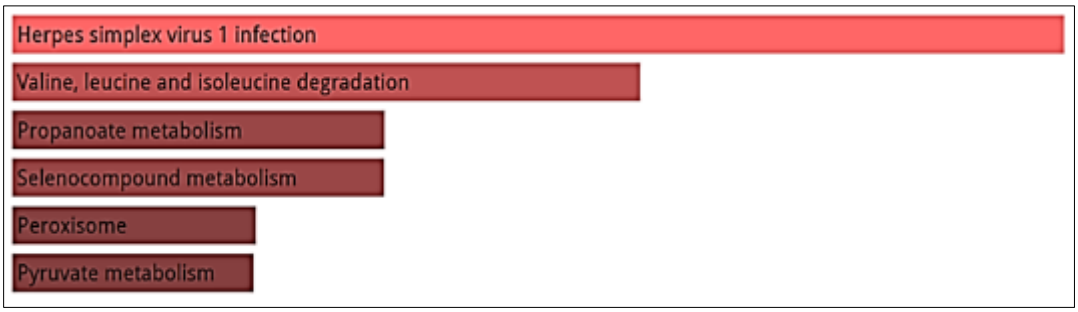


Figure 12 Top KEGG 2021 pathways overrepresented across differentially expressed genes for upregulated genes, detected by the Enrichr server (p-value <0.01). The lighter the colour used for the pathway, the lower its p-value

To check the genes that are involved in these six pathways (Figure 13). The clustergram tool from the Enrichr website was used to obtain the top 40 upregulated genes that were enhanced in the top six pathways. The results showed top upregulated genes in the top six pathways including for example, RNASEL, ZNF141 and ZNF175 genes are the most significant for the Herpes simplex virus 1 infection pathway (Figure 13).

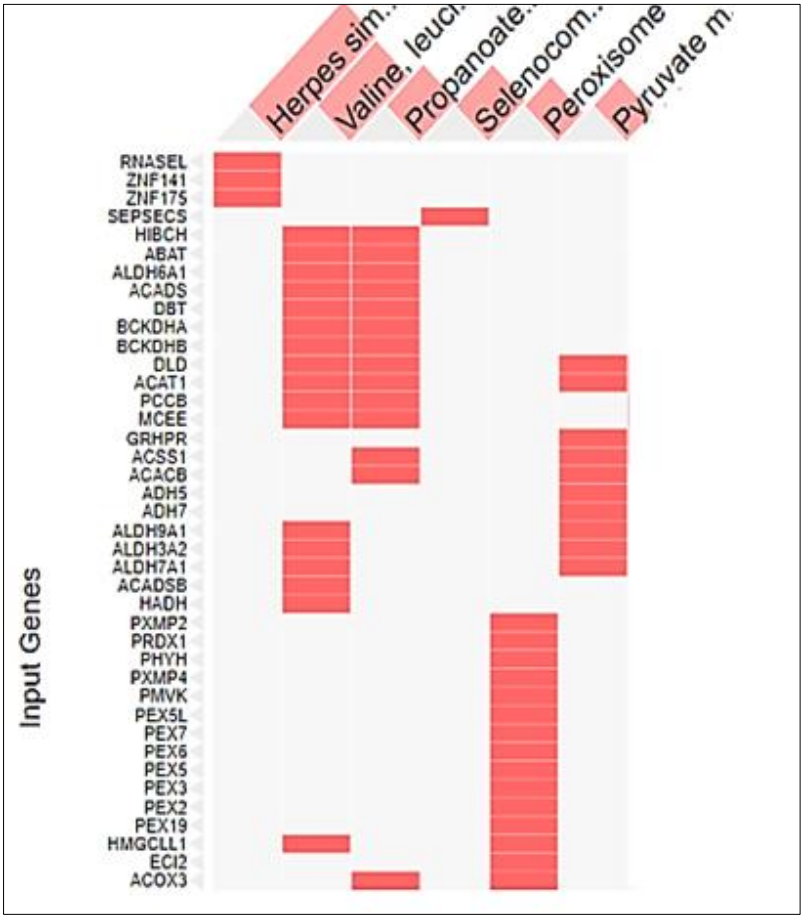


Figure 13 The clustergram diagram showing upregulated genes involved in 6 pathways in KTCN corneas. It detected through the Enrichr server (p-value <0.01). The rows represent genes. The columns represent pathways. The pink boxes show genes that have a role in the pathways

To identify the properties of the gene products for all 4,444 upregulated genes, the ontology tool in the Enrichr server was used to check the Gene Ontology (GO) cellular component, GO molecular function and GO biological process. The results show that peroxisome is shared across the three GO domains (Figure 14).

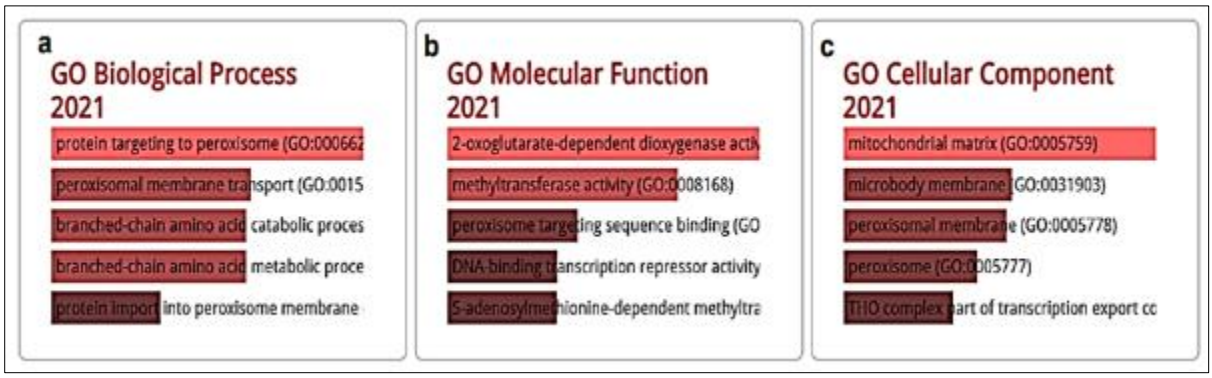


Figure 14 The Gene Ontology. It shows ontology domains GO biological process (a), GO molecular function (b), and GO cellular component (c), detected by the Enrichr server (p-value <0.01). Protein targeting to peroxisome has the most GO biological processes for KTCN from the upregulated genes. 2-oxoglutarate-dependent dioxygenase activity is the most prominent GO molecular function in the upregulated genes for KTCN. The mitochondrial matrix is the most prominent GO cellular component for KTCN in the upregulated genes. The lighter the colour, the lower p-value for the function

The Diseases/Drugs tool on the Enrichr website was used to check diseases related to these 4,444 upregulated genes, and this identified 8 diseases related to these genes including:

Elevated levels of phytanic acid, Zellweger Spectrum, and Profound global developmental delay and 5 other diseases have the lower p-value <0.01 than other disease (Figure 15).

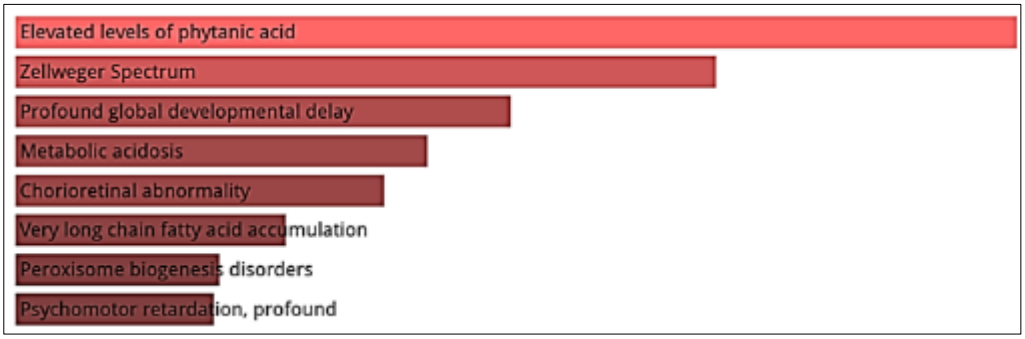


Figure 15 Diseases caused by the identified upregulated genes as detected by the Enrichr server (p-value <0.01). The diseases in the graph are the most common diseases for the upregulated genes. Elevated levels of phytanic acid has the lowest p-value, followed by Zellweger Spectrum disease. The lighter the colour used for the disease, the lower its p-value

To check which genes are involved in these diseases because this research focused on the upregulated genes (Figure 15), a clustergram was formed for the top 20 upregulated genes involved in the top eight diseases. The results showed that peroxisome biogenesis (PEX) genes are the most significant cause of phytanic acid, Zellweger spectrum disease, and profound global developmental delay (Figure 16).

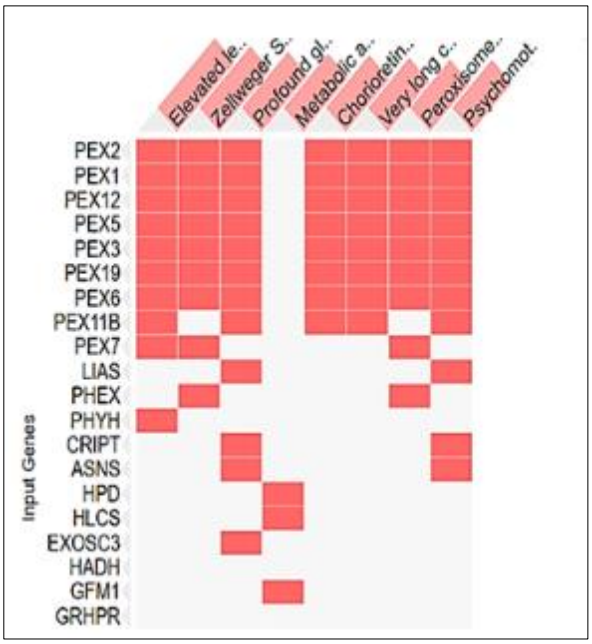


Figure 16 The cluster gram diagram showing upregulated genes involved in 8 diseases. It detected through the Enrichr server (p-value <0.01). The rows represent genes. The columns represent diseases. The pink boxes show genes that have a role in the pathways

4. Discussion

The RNA samples used in this research had already been assessed in an RNA-seq experiment (Kabza et al., 2017). These fifty RNA samples were derived from two independent RNA-Seq experiments: discovery RNA-Seq analysis and replication RNA-seq experiment (Kabza et al., 2017). Sixteen samples were from the discovery RNA-Seq analysis (eight KTCN and eight non-KTCN), sequenced with high-coverage RNA sequencing of 150-250 million read pairs per sample, which helps to detect novel transcripts in the samples examined (Kabza et al., 2017). Twenty-four samples were derived from the replication RNA-Seq experiment (seventeen KTCN and seventeen non-KTCN), sequenced with 50-70 million read pairs per sample (Kabza et al., 2017). This research analysed samples from both discovery RNA-Seq analysis and replication RNASeq experiment. However, the published study analysed only the replication study for detection of differentially expressed genes in cases and controls. This research analysed the data in different ways from those used in the published paper. This study used the vst function to normalise the data and the Enrichr server for pathway analysis, while the published paper used the edgeR packages to normalise the data and the ConsensusPathDB server for pathway analysis, and this difference led to slightly different results from those obtained in the published study (Kabza et al., 2017).

This study tested the relation between development of KTCN and certain factors which had not been examined previously. In addition, this study focused on upregulated genes and found a relation between defects in peroxisomes and development of KTCN, which is novel.

In this study, principal component analysis showed that five non-KTCN samples (KR_53, KR_57, KR_64, KR_08 and KR_35) were clustered with the KTCN cluster. This misclassification could be due to the heterogeneity of the control group, and had also been observed and detected as a study limitation for the original study using these samples (Kabza et al., 2017). Because it is difficult to obtain healthy corneas from healthy living individuals, control corneal tissues were obtained from individuals with diseases that lead to corneal scars (Kabza et al., 2017). Most of the samples that were misclustered (KR_53, KR_57, KR_64, KR_08 and KR_35) had been obtained from individuals with a shared corneal scar, ocular diseases such as bullous keratopathy, history of ocular trauma and corneal ulcer, and another disease. The ocular disease or corneal scaring in these misclassified samples might have caused changes in molecular pathways, leading these cases to share molecular pathways with KTCN in terms of inhibition or activation of those pathways.

This study determined the relation between certain factors: age, eye rubbing, and sex and developing KTCN. Regarding the age which is the age when the sample is taken, the PCA plots showed that Keratoconus is more prevalent in young adults and late teenagers. The first relevant study, on Collaborative Longitudinal Evaluation of Keratoconus (CLEK)

patients was conducted with 1,166 individuals in 2004, and shows that Keratoconus has a greater effect on people in early adulthood (Kymes et al., 2004). This study also showed that young adults might be more affected with keratoconus than older people Considering the relation between KTCN and eye rubbing, a PCA plot showed a significant correlation between this factor and the disease. In contrast, two studies find that there is no correlation between eye rubbing and KTCN. In a study in Jerusalem, it was found that there was no significant association between eye rubbing and KTCN (Millodot et al., 2011), while the second study was conducted with ninety-two individuals within the Lebanese student population in Lebanon, and showed that 12% of KTCN individuals had a family history, whereas eye rubbing was not a risk factor for developing KTCN (Van den Brink et al., 2012). However, a survey in 2013 which included 244 KTCN individuals found that approximately 65.6% of the sample for individuals rubbed their eyes (Shneor et al., 2013). A Saudi study showed that eye rubbing was the most widespread risk factor between KTCN individuals, and found that roughly 48% of KTCN patients reported eye-rubbing (AlShammari et al., 2016). This suggests that eye rubbing might be a risk factor in developing KTCN.

The PCA plot related to sex factor showed that males develop KTCN more than females. In 2003, a study from New Zealand was conducted with a New Zealand population and showed that males were affected more frequently and rapidly with keratoconus than females. In 2005, a study showed that females were affected more than males by ocular symptoms and had a family history of keratoconus greater than for males (Fink et al., 2005). Another study performed in Mexico among teenagers found that keratoconus affected 66% of females and 33.3% of males (Valdez-García et al., 2014). However, more recently, another study obtained different results to those of the 2003 and 2005 studies. The newer study showed that men, African-Americans and Latinos, and individuals with diseases such as asthma, sleep apnoea and Down's syndrome were more likely to develop keratoconus than others (Woodward et al., 2016, Lin, et al,. 2023). In this study, it was found that males were more affected than females, which might mean that Polish men are more likely to develop keratoconus, as with black and Latino persons. It was also noted that a larger proportion of the males studied had asthma and allergies as in the Appendix 1, which could contribute to the development of KTCN in males.

In this research, the hierarchical cluster analysis and heatmap of RNA-seq for gene expression data perfectly classified KTCN and non-KTCN samples after removing the misclassified samples. In addition, it shows that genes that are upregulated in KTCN samples are downregulated in non-KTCN samples, and that genes that are downregulated in KTCN samples are upregulated in non-KTCN samples. The research aimed to identify differential expressed genes. Downregulated genes showed significant dysregulation of 95 pathways in KTCN, including cytokine-cytokine receptor interaction, coronavirus disease, and viral protein interaction with cytokine and cytokine receptor. This study observed a decrease in the expression of cytokine genes, including IL6, TNF, IL1B, TGFB1, TGFB2 and TGFB3. One of the previously conducted studies showed that three TGF-β isoforms (TGFB1, TGFB2 and TGFB3) have a role in regulating the formation of the extracellular matrix (ECM) in tissues including the eye (Priyadarsini et al., 2015). Decreased expression of cytokine TGF-β contributes to formation of an abnormal ECM structure, leading to alterations in the cornea's structure by forming a cone shape in KTCN individuals (Priyadarsini et al., 2015).

The current study confirms that deregulated cytokine genes have a role in developing KTCN, by forming a pathological ECM structure. In addition, low levels of cytokine genes in KTCN corneas might cause a difficulty in fighting infections. However, cytokine level has been reported to be increased in the tears of KTCN Individuals, leading to cornea disruption (Jun et al., 2011). Overexpression of some cytokine genes such as IL-17 cause thinning and weakening of the cornea tissue (Jun et al., 2011).

The correlation between the downregulating cytokine genes and developing KTCN found in the published paper (Kabza et al., 2017). Pathway analysis showed that only six pathways from KEGG databases in the upregulated gene set were statistically significantly overrepresented. These include the RNASEL, ZNF141 and ZNF175 genes, which function in the Herpes simplex virus 1 (HSV_1) infection pathway and have a role in responding defensively against the virus. Those individuals who have upregulated RNASEL, ZNF141 and ZNF175 genes could have Herpes simplex Keratitis (HSK). This is a disease caused by an infection of the cornea with HSV-1, leading to damage to the eye, scarring and thinning of the cornea, or blindness (Liesegang, 2001). This could represent a relation between HSV-1 and cornea changes in the KTCN patients. It was also identified that some upregulated genes are involved in elevating phytanic acid and Zellweger Spectrum disease, including PEX1, PEX2, PEX3, PEX5, PEX6, PEX7, and PEX19, are involved in the peroxisome pathway and have a role in peroxisome cells, as shown in the GO domains: biological processes, molecular function, and cellular components. Defects on these genes (PEX1, PEX2, PEX3, PEX5, PEX6, PEX7, and PEX19) lead to defection in peroxisome function resulting in peroxisome biogenesis disorders (PBDs) (Steinberg et al., 2006). When the PEX7 gene is mutated or upregulated, as found here in this study, it leads to an increased level of phytanic acid in the body in some patients. An accumulation of phytanic acid as a toxin in the blood and tissues might lead to Refsum disease (RD), which is a slow, progressive disorder of the lipid metabolism (Van den Brink et al., 2012). These errors in the lipid metabolisms could

prevent the myelin sheath from surrounding nerve cells in the eye leading to disorders such as Retinitis Pigmentosa (RP). The affected retina in individuals with RP can become worse and develop Leber congenital amaurosis (LCA), which is a severe form of RP. LCA individuals develop an abnormal cornea, leading to formation of a cone shape to the front of the eye, and develop keratoconus (Damji et al., 2001). When genes PEX1, PEX2, PEX3, PEX5, PEX6 and PEX19 are mutated, they may cause Zellweger syndrome. These genes are involved in the peroxisome pathway and are required for biological processes, molecular function and cellular components for the peroxisome (Fujiki, 2016). Mutation in these genes leads to defects in the function of peroxisome and the accumulation of very long chain fatty acids (VLCFA), resulting in development of Zellweger syndrome (Fujiki, 2016). Zellweger syndrome and accumulation of VLCFA cause cornea issues which might lead to develop KTCN in some cases. In summary, despite using published data, the analysis techniques used, and results obtained in the current study are mostly different from the analysis and results in the published paper (Kabza et al., 2017). It is confirmed here that factors including age, eye rubbing, and sex might have a role in KTCN development. In addition, it is confirmed that the TGF-β pathway has a role in developing KTCN. In addition, this research study is the first report to show the relation between the expressed upregulated genes for peroxisomes and development of KTCN. Disruption of the peroxisome pathway might be a new vision into the aetiology of KTCN. Further studies could involve polymerase chain reaction (PCR) needing to take place to find the expression level of peroxisomes genes in KTCN cases in comparison with non-KTCN individuals.

5. Conclusion

Summarizing outcome of the study revealed that upregulated genes were enriched in six pathways, notably the peroxisome pathway. A potential association was identified between peroxisome dysfunction and the development of keratoconus. The overexpression of peroxisome-related genes was found to elevate phytanic acid levels, potentially leading to Zellweger spectrum disorders, which affect the cornea and may contribute to the onset of keratoconus. Overall, the study's findings have the potential to pave the way for the development of new treatments, ultimately benefiting individuals affected and society as a whole.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] ABU-AMERO, K. K., AL-MUAMMAR, A. M. & KONDKAR, A. A. 2014. Genetics of Keratoconus: Where Do We Stand? Journal of Ophthalmology, 641708.

[2] ALSHAMMARI, Z., ALSHAMMARI, R., ALORF, S., ALSHAMMARI, R., ALSHAMMARI, W. & ALSHAMMARI, W. 2016. Prevalence, clinical features and associated factors of keratoconus patients attending Ophthalmology Department, King Khalid Hospital, Hail City, Saudi Arabia. EC Ophthalmology, 3, 388-400.

[3] BALASUBRAMANIAN, S. A., MOHAN, S., PYE, D. C. & WILLCOX, M. D. 2012. Proteases, proteolysis and inflammatory molecules in the tears of people with keratoconus. Acta Ophthalmol, 90, e303-9.

[4] BLIGHE K, Lun A. 2021. PCAtools: PCAtools: Everything Principal Components Analysis. R Package version 2.4.0, https://github.com/kevinblighe/PCAtools.

[5] BRANCATI, F., VALENTE, E. M., SARKOZY, A., FEHÈR, J., CASTORI, M., DEL DUCA, P., MINGARELLI, R., PIZZUTI, A. & DALLAPICCOLA, B. 2004. A locus for autosomal dominant keratoconus maps to human chromosome 3p14–q13. Journal of Medical Genetics, 41, 188-192.

[6] CARLSON M .2019. Genome wide annotation for Human. R package version 3.8.2.

[7] CHAN, E., CHONG, E. W., LINGHAM, G., STEVENSON, L. J., SANFILIPPO, P. G., HEWITT, A. W., MACKEY, D. A. & YAZAR, S. 2021. Prevalence of Keratoconus Based on Scheimpflug Imaging: The Raine Study. Ophthalmology, 128, 515-521.

[8] DAMJI, K. F., SOHOCKI, M. M., KHAN, R., GUPTA, S. K., RAHIM, M., LOYER, M., HUSSEIN, N., KARIM, N., LADAK, S. S., JAMAL, A., BULMAN, D. & KOENEKOOP, R. K. 2001. Leber's congenital amaurosis with anterior keratoconus in Pakistani families is caused by the Trp278X mutation in the AIPL1 gene on 17p. Can J Ophthalmol, 36, 252-9.

[9] DAVIS, S. & MELTZER, P. S. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics, 23, 1846-7.

[10] FINK, B. A., WAGNER, H., STEGER-MAY, K., ROSENSTIEL, C., ROEDIGER, T., MCMAHON, T. T., GORDON, M. O. & ZADNIK, K. 2005. Differences in Keratoconus as a Function of Gender. American Journal of Ophthalmology, 140, 459.1-12.

[11] FUJIKI, Y. 2016. Peroxisome biogenesis and human peroxisome-deficiency disorders. Proceedings of the Japan Academy. Series B, Physical and biological sciences, 92, 463-477.

[12] GAJECKA, M., RADHAKRISHNA, U., WINTERS, D., NATH, S. K., RYDZANICZ, M., RATNAMALA, U., EWING, K., MOLINARI, A., PITARQUE, J. A., LEE, K., LEAL, S. M. & BEJJANI, B. A. 2009. Localization of a Gene for Keratoconus to a 5.6-Mb Interval on 13q32. Investigative Ophthalmology & Visual Science, 50, 1531-1539.

[13] GODEFROOIJ, D. A., DE WIT, G. A., UITERWAAL, C. S., IMHOF, S. M. & WISSE, R. P. 2017. Age-specific Incidence and Prevalence of Keratoconus: A Nationwide Registration Study. Am J Ophthalmol, 175, 169-172.

[14] GORDON-SHAAG, A., MILLODOT, M., SHNEOR, E. & LIU, Y. 2015. The genetic and environmental factors for keratoconus. BioMed research international, 2015, 795738- 795738.

[15] HAO, X.-D., CHEN, P., ZHANG, Y.-Y., LI, S.-X., SHI, W.-Y. & GAO, H. 2017. De novo mutations of TUBA3D are associated with keratoconus. Scientific Reports, 7, 13570-13570.

[16] HARDCASTLE, A. J., LISKOVA, P., BYKHOVSKAYA, Y., MCCOMISH, B. J., DAVIDSON, A. E., INGLEHEARN, C. F. 2021. A multi-ethnic genome-wide association study implicates collagen matrix integrity and cell differentiation pathways in keratoconus. Communications biology, 4, 266-266.

[17] HASHEMI, H., KHABAZKHOOB, M. & FOTOUHI, A. 2013. Topographic Keratoconus is not Rare in an Iranian population: The Tehran Eye Study. Ophthalmic Epidemiol, 20, 385-91.

[18] HÉON, E., GREENBERG, A., KOPP, K. K., ROOTMAN, D., VINCENT, A. L., BILLINGSLEY, G., et al. 2002. VSX1: A gene for posterior polymorphous dystrophy and keratoconus. Human Molecular Genetics, 11, 1029-1036.

[19] HOFSTETTER, H. W. 1959. A keratoscopic survey of 13,395 eyes. Am J Optom Arch Am Acad Optom, 36, 3-11.

[20] HUGHES, A. E., DASH, D. P., JACKSON, A. J., FRAZER, D. G. & SILVESTRI, G. 2003. Familial keratoconus with cataract: linkage to the long arm of chromosome 15 and exclusion of candidate genes. Investigative ophthalmology & visual science, 44, 5063-5066.

[21] HUTCHINGS, H., GINISTY, H., LE GALLO, M., LEVY, D., STOËSSER, F., ROULAND, J. F., ARNÉ, J. L., LALAUX, M. H., CALVAS, P., ROTH, M. P., HOVNANIAN, A. & MALECAZE, F. 2005. Identification of a new locus for isolated familial keratoconus at 2p24. Journal of Medical Genetics, 42, 88-94.

[22] JUN, A. S., COPE, L., SPECK, C., FENG, X., LEE, S., MENG, H., HAMAD, A. & CHAKRAVARTI, S. 2011. Subnormal cytokine profile in the tear fluid of keratoconus patients. PLoS One, 6, e16437.

[23] KABZA, M., KAROLAK, J. A., RYDZANICZ, M., SZCZEŚNIAK, M. W., NOWAK, D. M., GINTER-MATUSZEWSKA, B., POLAKOWSKI, P., PLOSKI, R., SZAFLIK, J. P. & GAJECKA, M. 2017. Collagen synthesis disruption and downregulation of core elements of TGF-β, Hippo, and Wnt pathways in keratoconus corneas. European Journal of Human Genetics, 25, 582-590.

[24] KHALED, M. L., HELWA, I., DREWRY, M., SEREMWE, M., ESTES, A. & LIU, Y. 2017. Molecular and Histopathological Changes Associated with Keratoconus. BioMed Research International, 2017, 7803029.

[25] Kondkar AA, Azad TA, Sultan T, Khatlani T, Alshehri AA, Lobo GP, Kalantan H, Al-Obeidan SA, Al-Muammar AM. 2023. Association between Polymorphism rs61876744 in PNPLA2 Gene and Keratoconus in a Saudi Cohort. Genes (Basel). 21;14(12):2108.

[26] KOLDE R. 2019. pheatmap: Pretty Heatmaps. R package version 1.0.12. https://CRAN.Rproject.org/package=pheatmap

[27] KRACHMER, J. H., FEDER, R. S. & BELIN, M. W. 1984. Keratoconus and related noninflammatory corneal thinning disorders. Survey of ophthalmology, 28, 293-322.

[28] KYMES, S. M., WALLINE, J. J., ZADNIK, K. & GORDON, M. O. 2004. Quality of life in keratoconus. Am J Ophthalmol, 138, 527-35.

[29] LI, X., RABINOWITZ, Y. S., TANG, Y. G., PICORNELL, Y., TAYLOR, K. D., HU, M. & YANG, H. 2006. Two-Stage Genome-Wide Linkage Scan in Keratoconus Sib Pair Families. Investigative Ophthalmology & Visual Science, 47, 3791-3795.

[30] LIESEGANG, T. J. 2001. Herpes simplex virus epidemiology and ocular importance. Cornea, 20, 1-13.

[31] Lin Q, Wang X, Han T, Zhou X. 2023. Identification of genetic variants in two families with Keratoconus. BMC Med Genomics. 21;16(1):299.

[32] LISKOVA, P., HYSI, P. G., WASEEM, N., EBENEZER, N. D., BHATTACHARYA, S. S. & TUFT, S. J. 2010. Evidence for Keratoconus Susceptibility Locus on Chromosome 14: A Genome-wide Linkage Screen Using Single-Nucleotide Polymorphism Markers. Archives of Ophthalmology, 128, 1191-1195.

[33] LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15, 550.

[34] MILLODOT, M., SHNEOR, E., ALBOU, S., ATLANI, E. & GORDON-SHAAG, A. 2011. Prevalence and associated factors of keratoconus in Jerusalem: a cross-sectional study. Ophthalmic Epidemiol, 18, 91-7.

[35] PAPALI'I-CURTIN, A. T., COX, R., MA, T., WOODS, L., COVELLO, A. & HALL, R. C. 2019. Keratoconus Prevalence Among High School Students in New Zealand. Cornea, 38, 1382-1389.

[36] PRIYADARSINI, S., MCKAY, T. B., SARKER-NAG, A. & KARAMICHOS, D. 2015. Keratoconus in vitro and the key players of the TGF-β pathway. Molecular vision, 21, 577-588.

[37] SHNEOR, E., MILLODOT, M., BLUMBERG, S., ORTENBERG, I., BEHRMAN, S. & GORDONSHAAG, A. 2013. Characteristics of 244 patients with keratoconus seen in an optometric contact lens practice. Clin Exp Optom, 96, 219-24.

[38] SONG T, SONG J, LI J, BEN HILAL H, LI X, FENG P, CHEN W. 2024. The candidate proteins associated with keratoconus: A meta-analysis and bioinformatic analysis. PLoS One. 14;19(3): e0299739.

[39] STEINBERG, S. J., DODT, G., RAYMOND, G. V., BRAVERMAN, N. E., MOSER, A. B. & MOSER, H. W. 2006. Peroxisome biogenesis disorders. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1763, 1733-1748.

[40] TANG, Y. G., RABINOWITZ, Y. S., TAYLOR, K. D., LI, X., HU, M., PICORNELL, Y. & YANG, H. 2005. Genomewide linkage scan in a multigeneration Caucasian pedigree identifies a novel locus for keratoconus on chromosome 5q14.3-q21.1. Genetics in Medicine, 7, 397-405.

[41] TUFT, S. J., HASSAN, H., GEORGE, S., FRAZER, D. G., WILLOUGHBY, C. E. & LISKOVA, P. 2012. Keratoconus in 18 pairs of twins. Acta Ophthalmol, 90, e482-6.

[42] TYYNISMAA, H., SISTONEN, P., TUUPANEN, S., TERVO, T., DAMMERT, A., LATVALA, T. & ALITALO, T. 2002. A Locus for Autosomal Dominant Keratoconus: Linkage to 16q22.3-q23.1 in Finnish Families. Investigative Ophthalmology & Visual Science, 43, 3160-3164.

[43] VALDEZ-GARCÍA, J. E., SEPÚLVEDA, R., SALAZAR-MARTÍNEZ, J. J. & LOZANO-RAMÍREZ, J. F. 2014. Prevalence of keratoconus in an adolescent population. Revista Mexicana de Oftalmología, 88, 95-98.

[44] VAN DEN BRINK, D. M., BRITES, P., HAASJES, J., WIERZBICKI, A. S., MITCHELL, J., LAMBERT-HAMILL, M., DE BELLEROCHE, J., JANSEN, G. A., WATERHAM, H. R. & WANDERS, WAKED, N., FAYAD, A. M., FADLALLAH, A. & EL RAMI, H. 2012. [Keratoconus screening in a Lebanese students' population]. J Fr Ophtalmol, 35, 23-9.

[45] WANG, Y., RABINOWITZ, Y., ROTTER, J. & YANG, H. 2000. Genetic epidemiological study of keratoconus: evidence for major gene determination. American journal of medical genetics, 93, 403-409.

[46] WICKHAM H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4.

[47] WOODWARD, M. A., BLACHLEY, T. S. & STEIN, J. D. 2016. The Association Between Sociodemographic Factors, Common Systemic Diseases, and Keratoconus: An Analysis of a Nationwide Heath Care Claims Database. Ophthalmology, 123, 457-65.e2